Implied volatility sentiment

Félix, Luiz; Kräussl, Roman; Stork, Philip

published in
Quantitative Finance
2020

DOI (link to publisher)
10.1080/14697688.2019.1696018

document version
Publisher's PDF, also known as Version of record

document license
Article 25fa Dutch Copyright Act

Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: https://www.tandfonline.com/loi/rquf20

Implied volatility sentiment: a tale of two tails

Luiz Félix, Roman Kräussl & Philip Stork

To cite this article: Luiz Félix, Roman Kräussl & Philip Stork (2020) Implied volatility sentiment: a tale of two tails, Quantitative Finance, 20:5, 823-849, DOI: 10.1080/14697688.2019.1696018

To link to this article: https://doi.org/10.1080/14697688.2019.1696018

Quantitative Finance, 2020
Vol. 20, No. 5, 823–849, https://doi.org/10.1080/14697688.2019.1696018

Implied volatility sentiment: a tale of two tails

LUIZ FÉLIX*†‡, ROMAN KRÄUSSL§ and PHILIP STORK¶‡

†APG Asset Management, Gustav Mahlerplein 3, 1082 MS Amsterdam, Netherlands
‡School of Business and Economics, VU University Amsterdam, de Boelelaan 1105,1081 HV Amsterdam, Netherlands
§Luxembourg School of Finance, University of Luxembourg, 4, rue Albert Borschette, 1246 Luxembourg, Luxembourg
¶Tinbergen Institute Amsterdam, Gustav Mahlerplein 117, 1082 MS Amsterdam, Netherlands

(*Received 26 April 2019; accepted 15 November 2019; published online 29 January 2020*)

We propose a sentiment measure jointly derived from out-of-the-money index puts and single stock calls: implied volatility (IV-) sentiment. In contrast to implied correlations, our measure uses information from the tails of the risk-neutral densities from these two markets rather than across their entire moneyness structures. We find that IV-sentiment measure adds value over and above traditional factors in predicting the equity risk premium out-of-sample. Forecasting results are superior when constrained ensemble models are used vis-à-vis unregularized machine learning techniques. In a mean-reversion strategy, our IV-sentiment measure delivers economically significant results, with limited exposure to a set of cross-sectional equity factors, including Fama and French's five factors, the momentum factor and the low-volatility factor, and seems valuable in preventing momentum crashes. Our novel measure reflects overweight of tail events, which we interpret as a behavioral bias. However, we cannot rule out a risk-compensation rationale.

Keywords: Sentiment; Implied volatility; Equity-risk premium; Reversals; Predictability; Machine learning

JEL classification: G12, G14, G17

1. Introduction

End-users of out-of-sample (OTM) options tend to overweight tail events. This behavioral bias, suggested by Tversky's and Kahneman's (1992) cumulative prospect theory, is claimed to be present in the pricing of OTM index puts and in OTM single stock calls (Barberis and Huang 2008, Polkovnichenko and Zhao 2013). Within the index option market, the typical end-users of OTM puts are institutional investors, who use them to protect their large equity portfolios. Because institutional investors have large portfolios and hold a substantial part of the total market capitalization, OTM index puts are frequently in high demand and, as a result, overvalued. The reason for such richness of OTM puts goes back to the 1987 financial market crash. Bates (1991) and Jackwerth and Rubinstein (1996) argue that the implied distribution of equity market expected returns from index options changed considerably following the 1987 market crash. Their findings demonstrate that, since the crash, a large shift in market participants' demand for such instruments took place, evidenced by the probabilities implied by options prices and an increased volatility skew. Bates (2003) suggests that even models adjusted for stochastic volatility, stochastic interest rates, and random jumps do not fully explain the high level of OTM puts' implied volatilities (IV). Accordingly, Garleanu *et al.* (2009) argue that excessive IV from OTM puts cannot either be explained by option-pricing models that take such institutional investors' demand pressure into account.†

It has been claimed that OTM calls on single stocks are also systematically expensive (Barberis and Huang 2008, Boyer and Vorkink 2014). In line with that claim, Bollen and Whaley (2004) state that changes in the IV structure of single stock options across moneyness are driven by the net purchase of calls by individual investors, who are the typical end-user of these options. The literature provides several explanations for such strong buying pressure of OTM calls by retail investors. For example, Mitton and Vorkink (2007) and Barberis and Huang (2008) propose models in which investors have a clear preference for positive return skewness, or 'lottery ticket' type

*Corresponding author. Email: luizfffelix@yahoo.com

† It is important to disentangle the (equity) hedging behavior of institutional investor to their overall trading activity. Studies, such as Frijns *et al.* (2018), provide some evidence that institutional investors price stocks rationally, supporting the idea that the argued behavioral bias might be confined to institutional investors' portfolio insurance decisions.

of assets. In consequence, retail investors overpay for these leveraged securities, making OTM calls expensive and causing them to yield low forward returns. Cornell (2009) presents another behavioral explanation for the overpricing of single stock calls: because investors are overconfident in their stock-picking skills, they buy calls to get the most 'bang for the buck'. A related explanation for the structural overpricing of single stock calls is leverage aversion or leverage constraint (see Frazzini and Pedersen 2014): because investors are averse to borrowing (levering) or constrained to do so, they buy instruments with implicit leverage to achieve their return targets.

Beyond this literature that supports the link between institutional and individual investor trading activity and the structural overvaluation of OTM options, we argue that short-term trading dynamics also influence the pricing of OTM options. For instance, Han (2008) provides evidence that the index option's IV smirk is steeper when professional investors are bearish. He concludes that the steepness of the IV structure across moneyness relates to investors' sentiment. In the same line, Amin *et al.* (2004) argue that investors bid up the prices of put options after increases in stock market volatility and rising risk aversion, whereas such buying pressure wanes following positive momentum in equity markets. Mahani and Poteshman (2008) argue that trading in single stock call options around earnings announcements is speculative in nature and dominated by unsophisticated retail investors. Lakonishok *et al.* (2007) show evidence that long call prices increased substantially during bubble times (1990 and 2000) and that most of the single stock options' market activity consists of speculative directional call positions. Lemmon and Ni (2011) discuss that the demand for single stock options (dominated by speculative individual investors' trades) positively relates to sentiment. Li *et al.* (2018) suggest that calls are overvalued versus puts after stock price increases with the reverse being true after stock price decreases. Lastly, Polkovnichenko and Zhao (2013) suggest that time-variation in overweight of tail events derived from index put options might depend on sentiment, whereas Félix *et al.* (2019) provide evidence that such dynamics largely link to sentiment in single stock options.

The above studies suggest that OTM index puts and single stock calls are overpriced but that the valuation misalignment fluctuates considerably over time, caused by changes in investor sentiment. In this paper, we delve deeper into this issue and investigate how OTM options from index puts and single stock calls relate to forward returns and overweight of small probabilities (i.e. tail events).

The first contribution of our paper is to evaluate the information content of OTM index puts and single stock calls jointly, as a measure of sentiment. We assess the ability of this measure to predict forward equity returns and, more specifically, equity market reversals, defined as abrupt changes in the market direction.† Because we find IV skews to be strongly linked to overweight of small probabilities, we hypothesize that reversals may follow not only periods of extreme IV skews but also periods of excessive overweight of tails.‡

One characteristic of the literature that analyzes the informational content of IV skews is that it evaluates index puts' IV skews and single stock calls' IV skews completely separated from each other. As such, our second contribution is that we are, to the best of our knowledge, the first in the literature to use IV skews jointly extracted exclusively from OTM options from both the index and single stock option market as an indicator for investors' sentiment. Our sentiment measure, the so-called *IV-sentiment*, is calculated as the IV of OTM index puts minus the IV of OTM single stock calls.§ We conjecture that our *IV-sentiment* measure is an advance on the understanding of investors' sentiment because it captures the very distinct nature of these markets' two main categories of end-users: (1) IV from OTM puts captures institutional investors' willingness to pay for leverage to hedge their downside risk (portfolio insurance), as a measure of bearishness, whereas (2) IV from OTM single stock calls captures levering by individual investors for speculation on the upside ('lottery tickets' buying), as a measure of bullishness. Thus, a high level of *IV-sentiment* indicates bearish sentiment, as IV from index puts outpace the ones from single stock calls. In contrast, low levels of *IV-sentiment* indicate bullishness sentiment, as IV from single stock calls become high relative to the ones from index puts.

We find that our *IV-sentiment* measure predicts equity market reversals better than overweight of small probabilities itself. It also delivers positive risk-adjusted returns more consistently than the common Baker and Wurgler (2007) sentiment factor when evaluated via two trading strategies, a high-frequency and a low-frequency one. In univariate and multivariate predictive regression settings, our *IV-sentiment* measure improves the out-of-sample forecast ability of traditional equity risk-premium models. This result is likely due to the uniqueness of our *IV-sentiment* measure relative to traditional predictive factors, as well as caused by the imposition of some structure into our models (in the form of

† Reversals in the context of this paper are not to be confused with the, so-called, reversal (cross-sectional) strategy, i.e. a strategy that buys (sells) stocks with low (high) total returns over the past month. We focus on the overall equity market, rather than investigating single stocks.

‡ The literature on the link between the skew and the overall stock market is still incipient. Doran *et al.* (2007) test IV skews as a predictor of aggregate market returns, but they study one-day ahead returns to skews, and ignore any longer effects. Other studies on the conditionality of forward equity market returns to other volatility-type of measures are: Ang and Liu (2007) for realized variance, Bliss and Panigirtzoglou (2004) for risk-aversion implied by risk-neutral probability distribution function embedded in cross-sections of options, Bollerslev *et al.* (2009) for variance risk premium, Driessen *et al.* (2013) for option-implied correlations, Pollet and Wilson (2008) for historical correlations, for implied volatility indices Rubbaniy *et al.* (2014) and Vilkov and Xiao (2013) for the risk-neutral tail loss measure. Most of these studies document a short-term negative relation between risk measures and equity market movements.

§ We acknowledge that the implied correlation and the correlation risk premia measures of Driessen *et al.* (2013) and Buss *et al.* (2017) are also jointly extracted from the index and single stock option markets. Nevertheless, the implied correlation is calculated using the entire cross-section of strikes, whereas our measure focuses on OTM options, i.e. the tails of the implied distribution.

coefficient constraints). Once these models are constrained, forecast ensemble approaches largely outperform individual predictors and unregularized machine learning techniques in predicting the equity risk-premium in our data set. These findings indicate the usefulness of some economic structure in modelling of the equity risk premium and highlight the danger of overfitting when machine learning techniques are not properly used. Thus, the third contribution of our paper is to complement the literature on out-of-sample forecasting of the equity risk-premium (Campbell and Thompson 2008, Welch and Goyal 2008, Rapach *et al.* 2010) by suggesting a new predictor, the *IV-sentiment* measure. Concurrently, we reiterate earlier findings that constrained linear models remain a powerful tool to forecast equity returns.

A final contribution of our work is to reveal the ability of our *IV-sentiment* measure to improve time-series momentum, cross-sectional momentum and equity buy-and-hold investment strategies. Our sentiment-based strategy is uncorrelated to these strategies, also at the tails, for instance, when cross-sectional momentum crashes contemporaneously to market rebounds (Kent and Moskowitz 2016). Consequently, we document an increase in the informational content of such strategies when combined with the *IV-sentiment* strategy, especially for cross-sectional momentum. In line with this outcome, we also report that returns from a *IV-sentiment*-based strategy are poorly explained by widely used equity risk factors, such as Fama and French's five-factors, the momentum factor (*WML*) and the low-volatility factor (*BAB*). Hence, we propose that active equity managers could benefit from *IV-sentiment* by time-varying their exposure to the market *Beta*.

The remainder of this paper is organized as follows. Section 2 describes the data and the main methods employed in our empirical study. In Section 3 we test how our sentiment proxy based on OTM options from both the index and single stock markets relates to forward equity returns and equity factors. In Section 4, we explore the link between *IV-sentiment* and overweight of small probabilities suggested by the CPT model as well as linking it to the Baker and Wurgler (2007) sentiment factor. Section 5 provides robustness tests and Section 6 concludes.

2. *IV-sentiment* measure

To compute our *IV-sentiment* measure, we use S&P 500 index options' IV data and single stock weighted average IV data from the largest 100 stocks of the S&P 500 index. The IV data used comes from closing mid-option prices from January 2, 1998 to March 19, 2013 for fixed maturities for four moneyness levels, i.e. 80, 90, 110, and 120, at the three-, six- and twelve-month maturity both for index and single stock options. IV for at-the-money (ATM) option, i.e. 100 moneyness level, is not employed in the calculation of *IV-sentiment* but is also part of our data set as used in subsequent analysis. Equation (A8l) in Appendix A shows how the weighted average single stock IV are computed.

We apply the S&P 500 index weights normalized by the sum of stock weights for which IVs across all moneyness levels are available. Following the S&P 500 index methodology and the unavailability of IV information for every stock on all days in our sample, stock weights in this basket change on a daily basis. The sum of weights is, on average, 58 percent of the total S&P 500 index capitalization and it fluctuates from 46 to 65 percent.

Continuously compounded stock market returns are calculated throughout our analysis from the basket of stocks weighted with the same daily-varying loadings used for aggregating the IV data† For index options, we use the S&P 500 index prices to calculate continuously compounded stock market returns. Realized index returns and single stock returns are downloaded via Bloomberg.

Our proposed IV skew sentiment metric, the so-called *IV-sentiment*, is a combined measure of the index and single stock options markets. Thus, our measure differs from the standard IV skew measures as it uses information from the two markets jointly instead of only capturing information from one market at a time. Our measure also differs from implied correlations as it uses information from OTM options and not the full cross-moneyness options' structure. Our *IV-sentiment* measure is specified as follows:

$$IV\text{-}sentiment = OTMindexputIV_{\tau p}\text{-}OTMsinglestockcallIV_{\tau c}, \tag{1}$$

where the subscript $\tau = 1 \cdots 3$ indexes the different option-maturities used, p specifies the moneyness levels 80 and 90 percent from index put options, and c specifies the moneyness levels 110 and 120 percent from single stock call options. Thus, our sentiment measure is calculated as permutations of IVs from the three-, six- and twelve-month maturities, and four points in the moneyness (80, 90, 110, and 120 percent) level grid, where the absolute distance from the two moneyness levels used per sentiment measure and the ATM level (100 percent moneyness) is kept constant. In other words, the *IV-sentiment* metric produced is restricted to the 80 minus 120 percent and the 90 minus 110 percent measures, hereafter called the *IV-sentiment 90-110* and *IV-sentiment 80-120* measures. From the granular data set across different moneyness levels and maturities, we create six distinct skew-based measures of *IV-sentiment*. Using such a construction, our *IV-sentiment* measure jointly incorporates bearishness sentiment from institutional investors and bullishness sentiment from retails investors.

3. Predicting with *IV-sentiment*

3.1. *IV-sentiment high frequency strategy*

We begin our predictability tests of the IV-sentiment measure by implementing a high frequency (daily) trading rule with the aim to predict equity market reversals. Our hypothesis is that

† We thank Barclays Capital for providing the implied volatility data. Barclays Capital disclosure: '*Any analysis that utilizes any data of Barclays, including all opinions and/or hypotheses therein, is solely the opinion of the author and not of Barclays. Barclays has not sponsored, approved or otherwise been involved in the making or preparation of this Report, nor in any analysis or conclusions presented herein. Any use of any data of Barclays used herein is pursuant to a license.*'

when the *IV-sentiment* measure is significantly higher (lower) than its normal level, overweight of small probabilities is then extreme and likely to mean-revert in the subsequent periods in tandem with the underlying market. The trading strategy, thus, suggests to buy (sell) equities when there is excessive bearishness (excessive bullishness/complacency) indicated by high (low) level of *IV-sentiment*.

The strategy is tested via a pair-trading rule among long and short positions in the S&P 500 index and a USD cash return index. For simplicity, it is implemented as a purely directional strategy where positions are constant in size and *IV-sentiment* is normalized via a Z-score. The trading rule enters a five percent long equities position when the *IV-sentiment* is higher than a pre-specified threshold, for example, its historical two standard deviation. The trading rule closes the position, by entering into a full cash position, when the normalized *IV-sentiment* measure converges back to its average. Conversely, the rule enters a short equities position when the *IV-sentiment* is lower than its historical negative two standard deviation threshold and buys back a full cash position when it converges to its average. A five basis points trading cost is charged over the five percent position traded in equities. In order to avoid strategy overfitting, we (1) compute the Z-score using multiple look-back periods, and (2) use multiple threshold levels to configure excessive sentiment.† We evaluate these contrarian strategies on a volatility-adjusted basis using standard performance analytics such as the information ratio (IR), downside risk characteristics, and higher moments of returns. We compare these strategies to (1) other contrarian strategies that make use of IV volatilities, such as an IV skew-based strategy, a volatility risk premia (VRP) strategy, and an implied-correlation-based (IC) strategy,‡ (2) the equity market beta, i.e. the S&P 500 index, and (3) alternative beta strategies, such as writing put options, a 110-95 collar strategy, the G10 FX carry, equity cross-sectional momentum, and a time-series momentum strategy.§ We further evaluate such strategies by estimating the paired correlation coefficient between them, as well as tail and (distribution) higher-moment dependency statistics such as conditional co-crash (CCC) probabilities (see Appendix A.3) and co-skewness. Our back-test samples start in January 2, 1998 and end in December 4, 2015.¶

The boxplots of IRs obtained by our *IV-sentiment* strategies and other IV-based strategies are provided in figure 1. We see that the *IV-sentiment 90-110* strategy seems to perform better than the *IV-sentiment 80-120* strategy, as the IR means and dispersion of the former strategy dominate the ones for the latter. The average IR for the *IV-sentiment 90-110* strategy is positive for the three- and six-month option maturities but negative for the twelve-month. For the three- and six-month strategies, all one-standard deviation boxes for the IR lay in positive territory, suggesting that the *IV-sentiment 90-110* strategy is robust to changes in look-back and outer-threshold parameters. Further, the *IV-sentiment 90-110* is superior to single-market IV skew-based strategies for the three- and six-month maturities, but not for the twelve-month maturity. At the three-month maturity, the average IR and dispersion for the *IV-sentiment 90-110* strategy are similar to the ones for the VRP strategy. However, for the six- and twelve-month maturities, the VRP strategies dominate the *IV-sentiment 90-110* based on the average IR, despite larger dispersion for the six-month maturity strategy.

Figure 1 shows that the IC strategies seem to deliver relatively high and consistent IRs, especially when using 80 and 90 percent moneyness levels. At the three- and six-month maturities, the performance of IC strategies matches the performance of the *IV-sentiment 90-110* and VRP strategies. At the twelve-month horizon, the 80 and 90 percent IC strategies are superior to the *IV-sentiment 90-110* measure. Overall, the boxplots in figure 1 suggest that the *IV-sentiment 90-110* strategy is robust to changes in parameters but also that its performance is matched by other IV-based strategies. Table 1 Panel A provides performance analytics for the *IV-sentiment 90-110* strategy, as well as for alternative strategies.

We observe that the *IV-sentiment 90-110* strategy (using three-month option maturity‖) delivers returns (20 basis points) and risk-adjusted returns (0.29) that are superior to many of the other strategies compared, such as the S&P 500, the IV skew, the VRP, the IC, the 90-110 collar, the G10 FX carry, and the equity momentum. The only strategies that deliver equal or higher risk-adjusted returns than our *IV-sentiment 90-110* strategy are the time-series momentum and the put writing. The return skewness for our *IV-sentiment* strategy is positive (0.10) and above the average of the other strategies. A strategy that has surprisingly high skewed returns is the IC (0.43). The drawdown characteristics such as the maximum drawdown, the average recovery time, and the maximum daily drawdown of our *IV-sentiment* strategy are somewhat similar to the other IV-based strategies.

In the following, we combine our *IV-sentiment* strategy with a simple buy-and-hold of the S&P 500 index, a cross-sectional equity momentum, and a time-series momentum strategy, on a standalone basis. These combinations are done by weighting returns in a 50/50 percent proportion. Statistics for the strategies are presented in columns (11) and (13) of Panel A of table 1. We note that the combined strategies improve the IRs of these three strategies. The IR for the S&P 500 rises from 0.14 to 0.29, for the time-series momentum from 0.71 to 0.75 and by a staggering 0.20 points for the cross-sectional momentum strategy, from 0.14 to 0.34. The drawdown and skewness characteristics are also improved,

† We also test a percentile normalization and find results qualitatively similar to the use of Z-scores.

‡ A IC (or dispersion trading) strategy buys (sells) index options and sells (buys) single stock options, while delta hedging, to arbitrage price differences in these two volatility markets.

§ Strategy return series used are, respectively, the CBOE S&P 500 BuyWrite Index, the CBOE Investable Correlation Index, the S&P 500 index, CBOE put writing index, the CBOE 110-95 collar, the DB G10 FX carry index, the JPMorgan Equity Momentum index, and the Credit Suisse Managed Futures index.

¶ As our *IV-sentiment* measure requires less (cross-sectional) IV data than the *Delta minus Gamma spread* to be calculated, we can extend our sample, from March 19, 2013, until December 4, 2015.

‖ Results provided by tables 1–3 are all based on the three-month option maturity. The choice made for the three-month maturity is due to the higher robustness suggested by the *IV-sentiment*-based strategies reported in figure 1. Results for the six-month maturity are qualitatively the same as the three-month maturity ones. For the twelve-month maturity results for active strategies are poorer as reported by figure 1, though, results for table 3 for this maturity are similar to the ones for shorter maturities.

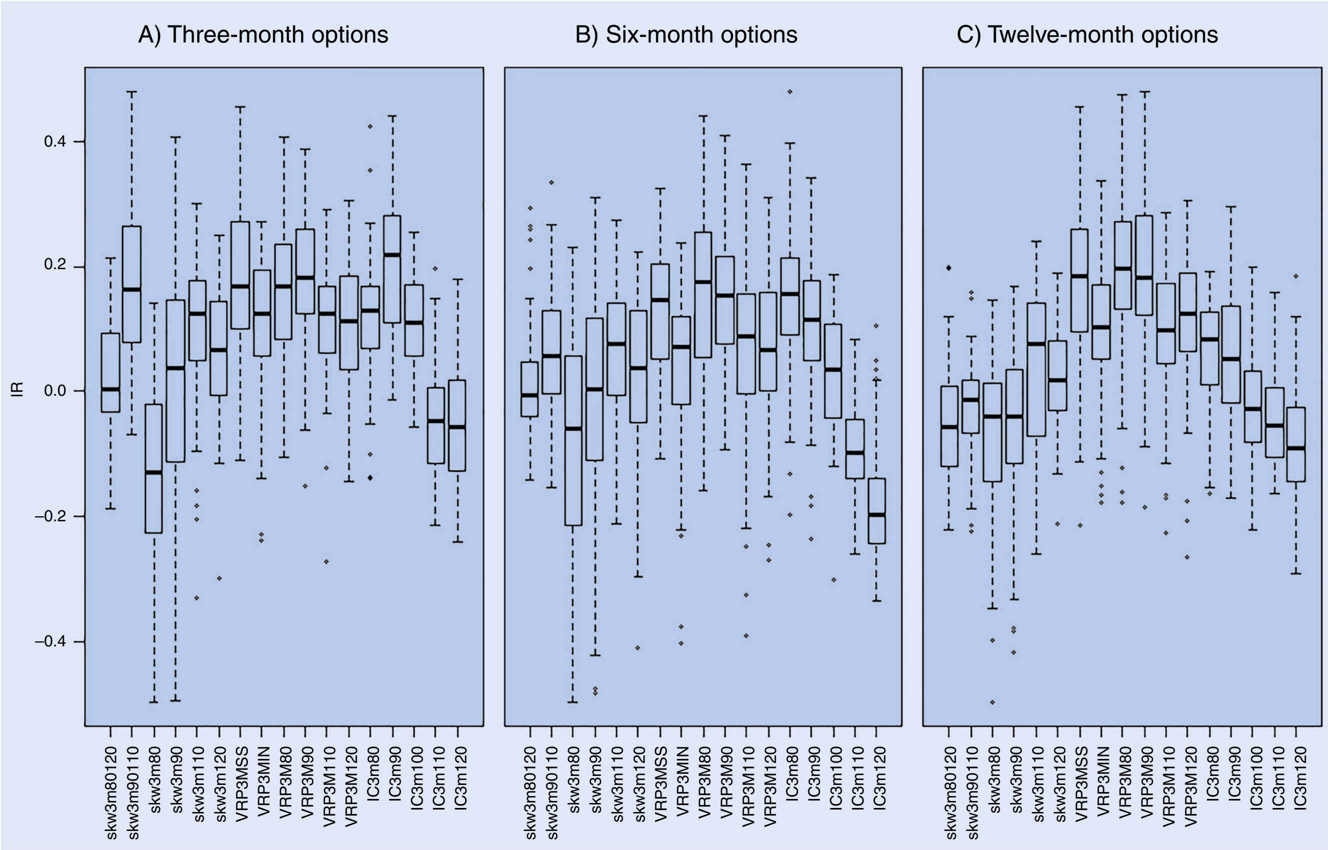

Figure 1. Information ratio boxplots for daily IV-based strategies. The boxplots depict the distribution of information ratios (IR) obtained by the IV-based strategies tested, when different look-back periods and outer-thresholds are used per factor-specific strategy. Boxplot A depicts the distribution of IRs when the IV factor used is obtained from three-month options. Panels B and C depict the same information while using the IV factors obtained from six- and twelve-month options, respectively. (a) Three-month options. (b) Six-month options. (c) Twelve-month options.

especially for the cross-sectional momentum strategy. We argue that these improvements in the IR and downside statistics occur due to the low correlation and low higher moments-/tail-dependencies of our *IV-sentiment* strategy with these alternative strategies. For instance, table 1 Panel B indicates that the *IV-sentiment* strategy is negatively correlated to both equity momentum and time-series momentum, by -0.16 and -0.11, respectively.

Co-skewness and, especially, CCC probabilities of the *IV-sentiment* strategy with momentum strategies are also very low (see Panel C of table 1). Since Kent and Moskowitz (2016) document that momentum occasionally crashes, in particular cross-sectional momentum, we suggest that the large improvement delivered by *IV-sentiment* to these strategies is likely due to the reduction of their large negative tails.

Moreover, table 1 Panel B indicates that the *IV-sentiment* strategy is, on average, positively related to other strategies. The highest correlation observed for the *IV-sentiment* strategy is with the IC strategy (0.70), which is an intuitive result given that these are the only two strategies driven jointly by the index option market and the single stock option market. The correlations of our *IV-sentiment* strategy with other IV-based strategies are also relatively high: 0.18 with the VRP and 0.41 with the IV skew 90 percent. The correlation of the *IV-sentiment* with the S&P 500 index is with 0.10, very low. The correlation of the *IV-sentiment* strategy with other strategies that perform poorly in ‘bad times’ is also low, at 0.04 with the put writing, at 0.07 with the G10 FX carry, and at 0.13 with the 90-110 collar strategy. We also note that some other strategies are highly correlated with each other, e.g. with 0.89 between the S&P 500 and the put-writing, whereas negative correlations are mostly observed for momentum strategies. Our findings on correlations among strategies are mostly reiterated by the estimated tail-dependence between them using co-skewness and CCC probabilities reported in Panel C of table 1.

As a robustness check, we analyze whether our *IV-sentiment* high-frequency trading strategy performs well due to both its legs or whether its merit is concentrated in either the long- or the short-leg. We separate the performance of the two legs of the strategy as if they were two different strategies and we compute individual performance statistics.

We find that the median IRs of long-legs are substantially higher than for short-legs.† The IR distributions of the short positions seem slightly skewed to the negative side, whereas for the long positions they seem skewed to the positive side. These results indicate that the merit of our *IV-sentiment* strategy is concentrated in its buy- rather than its sell-signal. This outcome also applies to other IV-based strategies which have

† These results are shown in figure 6 contained in the online Appendix C.1, available at https://github.com/luizfelix/IV-Sentiment.

Table 1. *IV-sentiment* based pair-trade strategy.

Panel A - Back-test results	(1) IV-sentiment 90-110	(2) IV Skew 3m 90	(3) VRP 3m 90	(4) IC 3m 90	(5) S&P500	(6) Put writing	(7) 110-95 collar	(8) G10 FX carry	(9) Equity Momentum	(10) CTA	11 S&P500 + IVSent	12 Equation Mom + IVSent	13 CTA + IVSent
Average return	0.20%	0.14%	0.12%	0.17%	0.10%	0.34%	0.14%	0.18%	0.10%	0.51%	0.21%	0.24%	0.53%
Volatility	0.71%	0.71%	0.71%	0.71%	0.71%	0.71%	0.71%	0.71%	0.71%	0.71%	0.71%	0.71%	0.71%
Information ratio	0.29	0.20	0.17	0.24	0.14	0.48	0.20	0.26	0.14	0.71	0.29	0.34	0.75
Skewness	0.10	− 0.07	− 0.01	0.43	− 0.18	− 0.60	0.01	− 0.93	− 0.45	− 0.37	− 0.05	0.09	− 0.37
Kurtosis	15.84	24.73	29.02	18.54	8.12	24.13	2.25	12.04	4.23	2.89	7.66	17.53	2.91
Max drawdown	− 1.7%	− 1.6%	− 2.9%	− 1.7%	− 3.0%	− 2.5%	− 2.9%	− 3.2%	− 2.9%	− 1.4%	− 2.4%	− 2.2%	− 1.1%
Avg recovery time (in years)	0.43	0.42	0.41	0.35	0.22	0.06	0.20	0.16	0.25	0.14	0.13	0.21	0.14
Max daily drawdown	− 0.55%	− 0.53%	− 0.49%	− 0.47%	− 0.34%	− 0.53%	− 0.29%	− 0.50%	− 0.35%	− 0.31%	− 0.29%	− 0.43%	− 0.48%

Panel B - Correlation matrix	IV-sentiment 90-110	IV Skew 3m 90	VRP 3m 90	IC 3m 90	S&P500	Put writing	110-95 collar	G10 FX carry	Equity Momentum	CTA
IV-sentiment	1	0.41	0.18	0.70	0.10	0.04	0.13	0.07	− 0.16	− 0.11
IV Skew	0.41	1	0.55	0.59	0.18	0.16	0.08	0.16	− 0.05	− 0.03
VRP	0.18	0.55	1	0.51	0.41	0.42	0.18	0.15	− 0.13	− 0.11
IC	0.70	0.59	0.51	1	0.34	0.31	0.24	0.14	− 0.21	− 0.13
S&P500	0.10	0.18	0.41	0.34	1	0.89	0.88	0.28	− 0.05	− 0.15
Put writing	0.04	0.16	0.42	0.31	0.89	1	0.68	0.26	− 0.08	− 0.16
110-95 collar	0.13	0.08	0.18	0.24	0.88	0.68	1	0.23	0.13	− 0.05
G10 FX carry	0.07	0.16	0.15	0.14	0.28	0.26	0.23	1	0.02	− 0.05
Equity Momentum	− 0.16	− 0.05	− 0.13	− 0.21	− 0.05	− 0.08	0.13	0.02	1	0.30
CTA	− 0.11	− 0.03	− 0.11	− 0.13	− 0.15	− 0.16	− 0.05	− 0.05	0.30	1

Panel C - Tail dependence with IV-sentiment	IV-sentiment 90-110	IV Skew 3m 90	VRP 3m 90	IC 3m 90	S&P500	Put writing	110-95 collar	G10 FX carry	Equity Momentum	CTA
Co-skewness	1.6E − 12	− 2.9E − 13	1.0E − 11	4.6E − 12	− 6.5E − 10	− 3.3E − 09	9.6E − 10	− 9.8E − 10	5.9E − 10	− 3.2E − 09
1% cond. crash prob.	100%	51%	36%	77%	23%	21%	19%	9%	19%	2%
2% cond. crash prob.	100%	46%	37%	76%	32%	26%	26%	13%	15%	2%
5% cond. crash prob.	100%	44%	37%	78%	29%	32%	26%	17%	18%	7%

Panel A reports the results of contrarian pair-trade strategies based on our *IV-sentiment 90-110* indicator and on other IV-based strategies such as the IV Skew, the volatility-risk premia (VRP), and other traditional and alternative beta strategies, i.e. buy & hold the S&P500 index, put writing, 110-95 collar, G10 FX carry, cross-sectional equity momentum, and time-series momentum. The IV-based strategies use 252 days as the look-back period and $+/-$ two standard deviations as convergence thresholds. The columns (11) and (12) of Panel A report statistics for strategies that combine the three-month *IV-sentiment 90-110* strategy (column (1)) with the buy & hold the S&P 500 index (column (5)) and the time-series momentum strategy (column (10)). Panel B reports the correlation coefficients of daily returns estimated over the period between January 2, 1998 and December 4, 2015, for the same strategies reported in Panel A. Panel C reports the co-skewness and the conditional co-crash (CCC) probabilities of the three-month *IV-sentiment 90-110* with the other strategies, which indicate the degree of tail-dependence among them.

their long-legs outperforming their short-legs. This finding suggests that extreme bearish sentiment signals may be more reliable than extreme bullish sentiment signals.

Our results, thus, offer additional findings to the literature that explores the link between variance-measures and forward returns. Most of these studies recognize a negative and short-term relation between risk measures and returns, where a high variance links to subsequent negative to low returns. In contrast, our findings suggest that a high level of IV skew relates to subsequent positive and high returns. Our finding is mostly in line with Bollerslev *et al.* (2009), who document that equity market reversals are predicted by the variance risk-premium, but it also reiterates the conclusion of Xing *et al.* (2010), who suggest that equity markets are slow to incorporate the information embedded in implied volatility skews.

Further, we aimed to compare the trading performance of the Baker and Wurgler (2007) sentiment measure to our high-frequency strategy but this was not possible as the former factor is only available on a monthly or quarterly frequency and was only published until 2010. Thus, in a next step, we compare how trading strategies using our *IV-sentiment* measure compare to strategies that use the sentiment factor of Baker and Wurgler (2007). We do this by implementing a low-frequency pair trading strategy using both predictors. This pair-trading strategy is identical to the one applied above with the only difference being the rebalancing frequency and the number of observations in the look-back window. We use the following look-backs for the calculation of Z-scores: 1, 3, 6, 9, 12, 18, and 24 months. The *IV-sentiment* measures used are the *IV-sentiment 80-120* and *90-110* factors, available in our three different option maturities. Trading costs and strategy exits are the same as for the high-frequency pair-trade strategy. Figure 2 provides our results by a series of boxplots. The empirical findings are displayed in columns for the different option maturities and in rows for the different statistics evaluated: (1) IR, (2) return skewness, and (3) horizon, proxied by the average drawdown length (in months) observed per strategy.

Our findings suggest that the IRs of the *IV-sentiment* strategies are much less dispersed than the ones for the sentiment factor by Baker and Wurgler (2007). The median IR for the *IV-sentiment 90-110* factor is also higher than for the other two strategies. The *IV-sentiment 90-110* factor is the only strategy in which almost all backtests deliver positive IRs, with the exception of a few outliers. This is not the case for the other strategies, as a substantial amount of backtests deliver negative IRs. In line with our earlier results, the *IV-sentiment 90-110* factor dominates the *IV-sentiment 80-120* factor. The return skewness for the *IV-sentiment 90-110* strategy also dominates the ones for the other two strategies, as all boxplot features are superior. The *IV-sentiment 90-110* factor delivers the lowest median horizon of all strategies. The average horizons estimated for the *IV-sentiment 90-110* factor are 12, 13, and 19 months, respectively, for the strategies based on the three-, six- and twelve-month options. The dispersion of strategies' horizon is, however, higher for the *IV-sentiment 90-110* factor than for the Baker and Wurgler (2007) sentiment factor. We conclude that our *IV-sentiment* measure outperforms a trading strategy based on the sentiment factor by Baker and Wurgler (2007) on several key aspects: IR, return skewness, and trade horizon.

3.2. *Out-of-sample equity returns predictive tests*

3.2.1. Univariate models and ensemble forecast. Following our hypothesis that extreme bearishness and bullishness sentiment is likely followed by reversals in equity markets, we test in the following whether our *IV-sentiment* measure has out-of-sample predictive power in forecasting the equity risk premium, in line with the analysis of Welch and Goyal (2008). We follow the methodology used by Campbell and Thompson (2008) and Rapach *et al.* (2010), who build on Welch and Goyal (2008). Similarly to these three studies, our predictive OLS regressions are formulated as:

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \epsilon_{t+1}, \tag{2}$$

where r_{t+1} is the monthly excess return of the S&P 500 index over the risk-free interest rate, x_t is an explanatory variable hypothesized to have predictive power, and ϵ_{t+1} is the error term. Our predictive regressions also use the monthly data set provided by Welch and Goyal (2008),† but the scope of 14 explanatory variables used closely follows Rapach *et al.* (2010).‡

From the predictive regressions in equation (2), we generate out-of-sample forecasts for the next month ($t+1$) by using an expanding window. Following Rapach *et al.* (2010), the first parameters are estimated using data from 1947:1 until 1964:12, and forecasts are produced from 1965:1 until 2014:12. The estimating window for B/M starts slightly after 1947:1, while the number of observations available allows forecasting B/M to start also at 1965:1. For the *IV-sentiment*-based regression, the data used for the first parameter estimation starts at 1998:1 and ends at 1999:12 so that out-of-sample forecasting is performed from 2000:1 to 2014:12 only.

Following Campbell and Thompson (2008) and Rapach *et al.* (2010), restrictions on the regression model specified by equation (2) are applied. The first restriction concerns the sign of the slope coefficients of equation (2) for the 14 Welch and Goyal (2008) variables we employed. The second restriction comprises setting negative forecasts of the equity risk premium to zero. We specify an additional model containing both coefficient and forecast sign restrictions. The original equation (2) with no restrictions applied is called the *unrestricted model*, whereas the model with the two restrictions is called the *restricted model*. Once individual forecasts for r_{t+1} are obtained using the restricted and unrestricted models for every variable, weighted measures of central tendency (mean and median) of the N forecasts are generated by equation (3):

$$\hat{r}_{c,t+1} = \sum_{i=1}^{N} \omega_{i,t} \hat{r}_{i,t+1}, \tag{3}$$

† Welch and Goyal (2008) monthly data was updated until December 2014 and is available at http://www.hec.unil.ch/agoyal/.

‡ These variables are: the dividend price ratio, the dividend yield, the earnings-price ratio, the dividend-payout ratio, the book-to-market ratio, the net equity issuance, the Treasury bill rate, the long-term yield, the long-term return, the term spread, the default yield spread, the default return spread, the inflation rate, and the stock variance.

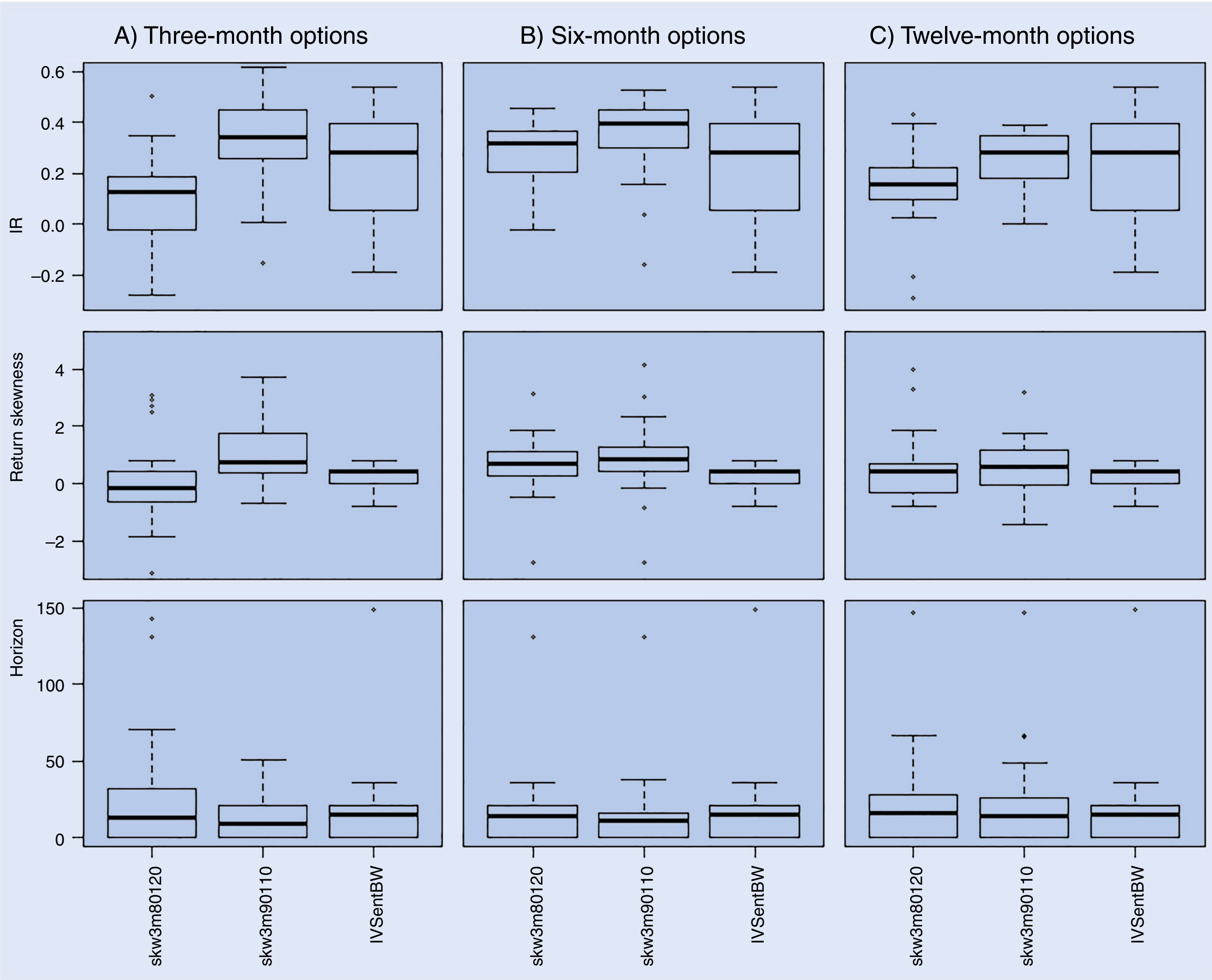

Figure 2. Information ratio, skewness and horizon for monthly IV-based strategies. The boxplots depict the distribution of information ratios (IRs), return skewness, and trade horizon (average drawdown) obtained by the *IV-sentiment* strategies tested, as well as the Baker and Wurgler (2007) sentiment factor when different look-back periods and outer-threshold are used per strategy. Boxplot A depicts the distribution of these statistics when the IV factor used is obtained from three-month options. Panel B and C depicts the same information but, respectively, when the IV factors used are obtained from six- and twelve-month options. Boxplots of IR, return skewness and trade horizon for the Baker and Wurgler (2007) factor are the same across option horizons but are shown for comparison with the *IV-sentiment* strategies. (a) Three-month options. (b) Six-month options. (c) Twelve-month options.

where $(\omega_{i,t})_{i=1}^{N}$ are the combining weights available at time t. Our forecast ensemble method is a more simple and agnostic approach than the one used by Rapach *et al.* (2010).† The mean and median ensemble methods are simply the equal weighed ($\omega_{i,t} = 1/N$) average and median of the forecasts. Our benchmark forecasting model is the historical average model with the use of an expanding window.

We use the out-of-sample R^2 statistic method (R^2_{OS}) introduced by Campbell and Thompson (2008) and followed by Rapach *et al.* (2010) for forecast evaluation. This method compares the performance of a return forecast $\hat{r}_{t+1}$ and a benchmark or naïve return forecast $\bar{r}_{t+1}$ with the actual realized return (r_{t+1}). We note that this method can be applied either to the single factor-based forecast models as well as to the ensemble or multifactor forecast models, both described in the previous section. The R^2_{OS} statistic is given as:

$$R^2_{OS} = 1 - \frac{\sum_{k=q_{0+1}}^{q} (r_{m+k} - \hat{r}_{m+l})^2}{\sum_{k=q_{0+1}}^{q} (r_{m+k} - \bar{r}_{m+l})^2}, \tag{4}$$

which evaluates the return forecasts from a predictive model (in the numerator) and the return forecasts from a benchmark or naïve model (in the denominator) by comparing the mean squared prediction errors (MSPE) for both methods. Because the ratio of MSPEs is subtracted from 1 in the R^2_{OS} statistic, its interpretation becomes: if $R^2_{OS} > 0$, then MSPE of $\hat{r}_{t+1}$ is smaller than for $\bar{r}_{t+1}$, indicating that the forecasting model outperforms the naïve (benchmark) model, and vice-versa.

† Rapach *et al.* (2010) classify their ensemble methods in two classes: the first class uses a mean, median, and trimmed mean approach. The second class uses a discounted mean square prediction error method, which combines weights as a function of the historical forecasting performance of the individual models during the out-of-sample period. This method weights more recent forecasts heavier than older ones by the use of one additional parameter. Despite the desirable features of such a second class combination method, we prefer to stick to the first class methods because they are more transparent and do not require the choice of an additional parameter.

To better evaluate the out-of-sample performance of models graphically, we employ the cumulative cum of squared error difference ($CSSED_{OS}$) statistic given below. The advantage of $CSSED_{OS}$ over R^2_{OS} is that it starts at zero and accumulates over time in a homoscedastic manner, whereas R^2_{OS} typically displays a very high volatility at the start of the (accumulation) period and a lower volatility of the metric as t increases†:

$$CSSED_{OS} = \sum_{k=q_{0+1}}^{q} (r_{m+k} - \bar{r}_{m+l})^2 - \sum_{k=q_{0+1}}^{q} (r_{m+k} - \hat{r}_{m+l})^2. \tag{5}$$

The results from our out-of-sample equity returns predictive tests are reported in table 2. Panel A reports the findings for the out-of-sample forecasting period between 1965:1 and 2014:12 for all individual variables except our *IV-sentiment* factor (*IV Sent*), for which forecasts are only available from 2004:1–2014:12, and for the ensemble forecasts. For individual models, R^2_{OS} comes from the restricted model, whereas for the aggregated models, the results are reported for both the restricted and the unrestricted models. The results of the aggregate models are reported in means and medians, reflecting the aggregation method used.

Panel A suggests that performance is not consistent across factors within the longer history of the out-of-sample test. Some factors outperform others by a large amount. Concurrently, the performance of most single factors is quite inconsistent through time, as figure 3 depicts: the slope and levels of $CSSED_{OS}$ constantly change from negative to positive and vice-versa for almost all factors. For some of them, $CSSED_{OS}$ even flips sign at times within the sample. In contrast, the aggregated models deliver better performance across restricted and unrestricted models using either averages or medians for aggregation method. Moreover, the performance of the weakest aggregate model (0.63) is superior to the best individual factor (*INFL* at 0.48) within the full sample.

Once we evaluate the 2004:1–2014:12 period, when *IV Sent* is used, we observe that the performance across factors remains inconsistent. The performance across individual factors looks less dispersed in this sample than in the full sample, but the overall performance deteriorates. The *IV Sent* factor performs well (ranging from 1.59 to 2.45 depending on the maturity), despite being strongly outperformed by the *SV AR* factor, while other factors perform extremely poorly (*NTIS* at −2.63, *INFL* at −2.58). The ensemble models that do not include *IV Sent* in their median versions (restricted and unrestricted) underperform the naïve forecasting benchmark as their R^2_{OS} is negative. Interestingly, when our *IV Sent* factor is added to these models, the performance improves substantially, outperforming the benchmark. We observe the same for models based on the mean: the mean-unconstrained and the mean-constrained models ex-*IV Sent* show a R^2_{OS} of 0.25 and 0.40, respectively. When the *IV Sent* factor is added to them, R^2_{OS} improves to 0.63 and 0.75, respectively. Therefore, our *IV Sent* factor seems to impact the ensemble model in a very distinct way when compared to other factors. R^2_{OS} from models that use median forecasts are worse than for models that aggregate forecasts by averaging. Nonetheless, improvements delivered by the inclusion of *IV Sent* and the imposition of model constraints are qualitatively the same across models aggregated by either median or averaging.

We also find that the correlation coefficient of the *IV Sent* using three-month options with other individual factors is most of the times negative or close to zero, and only exceeds 0.5 when evaluated against long-term yield (*LTY*).‡ Such correlation is higher for the *IV Sent* factor using six- and twelve-month option maturities. These results suggest that the improvements made by our *IV Sent* factor to the ensemble models stem partially from diversification benefits rather than from forecast performance (R^2_{OS}) alone.

3.2.2. ‘Kitchen sink’ and machine learning-based models. Further, we also test a ‘kitchen sink’ model§ as used by Welch and Goyal (2008) and Rapach *et al.* (2010) but we extend it toward machine learning algorithms. Our aim is to test whether more advanced models can fix the exceptionally poor out-of-sample performance of the multivariate approach to forecast the equity risk premium, as reported by Welch and Goyal (2008) and Rapach *et al.* (2010). The models tested by us in addition to the ‘kitchen sink’ OLS model are: (1) *Ridge regression*, (2) *Principal Component Regression*, (3) *Random forest*, and (4) *Deep Neural Networks*.¶ Our hypothesis for performing this models’ ‘horse race’ is that machine learning-based models might be able to improve over the multivariate OLS regression by either (1) reducing its variance and, so, avoiding overfitting, (2) better modelling potentially non-linearities present in the data, and (3) dampening the effect of collinearity in the regressors.

Our results from testing a ‘kitchen sink’ OLS model reiterate the ones of Welch and Goyal (2008) and Rapach *et al.* (2010) (see table 2). The model is the worst performing one in R^2_{OS} terms across all univariate and multivariate models. In contrast, individual machine learning algorithms using the same set of variables outperform the ‘kitchen sink’ model but do not consistently outperform the models that ensemble forecasts from univariate models. The *Ridge regression* model seems to be the best performing across all multivariate models as it delivers high R^2_{OS} in the 1965:1–2014:12 sample and a less negative R^2_{OS} than other models in the 2004:1–2014:12 sample. Given its linear character, the main advantages of *Ridge regression* over the ‘kitchen sink’ is the

† The undesirable graphical pattern of R^2_{OS} is caused by the normalization through $\sum_{k=q_{0+1}(r_{m+k}-\bar{r}_{m+l})^2}^{q}$, which at the start of the sample tends to be very small relative to $CSSED_{OS}$. Note that $R^2_{OS} = CSSED_{OS}/\sum_{k=q_{0+1}(r_{m+k}-\bar{r}_{m+l})^2}^{q}$.

‡ A full correlation matrix among the individual predictive factors tested by Rapach *et al.* (2010) and *IV-sentiment* factors can be provided upon request.

§ The ‘kitchen sink’ includes all 14 predictive variables used in our univariate models.

¶ We tune *Ridge regression* by using cross-validation with 10 folds. We tune our *Random forest* model using a single pass of out-of-bag errors to estimation of the optimal number of predictors sampled for spliting at each node. We use cross-validation in the estimation of our *Deep Neural Networks* model to come up with the number of layers and neurons (among a set of pre-defined structures) only. We do not apply any early-stopping procedure. A detailed description of these models and tuning procedures is out of scope of this paper. For specifics on these models, see Hastie *et al.* (2008)

Table 2. Out-of-sample equity risk premium.

Individual predictive regression model forecast				Ensemble forecasts		Machine learning methods	
Predictor	$R^2_{OS}(\%)$	Predictor	$R^2_{OS}(\%)$	Ensemble methods	$R^2_{OS}(\%)$	Methods	$R^2_{OS}(\%)$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. 1965:1–2014:12 out-of-sample period							
D/P	− 0.30	LTY	− 0.28	Mean-Unconstrained	1.08	Kitchen-sink (OLS)	− 88.14
D/Y	− 0.11	TMS	− 0.50	Median-Unconstrained	0.64	Ridge regression	0.81
E/P	− 0.41	LTR	0.22			Principal Component Regression	− 5.93
D/E	− 0.76	DFY	− 0.69	Mean-Constrained	1.11	Random Forest	− 9.97
B/M	− 0.88	DFR	− 0.55	Median-Constrained	0.63	Deep Neural Networks	− 84.14
NTIS	− 0.83	TBL	− 0.01			Mean-models	− 6.35
INFL	0.48					Median-models	− 2.06
SVAR	0.02						
Panel B. 2004:1–2014:12 out-of-sample period							
D/P	− 0.82	LTY	0.62	Mean-Unconstrained	0.25	Kitchen-sink (OLS)	− 62.64
D/Y	− 0.53	TMS	− 0.94	Median-Unconstrained	− 0.35	Ridge regression	− 1.76
E/P	− 1.31	LTR	0.01	Mean-Unconstrained + IVSent	0.63	Principal Component Regression	0.09
D/E	− 2.13	DFY	− 1.26	Median-Unconstrained + IVSent	0.27	Random Forest	− 8.80
B/M	− 0.16	DFR	− 0.64	Mean-Constrained	0.40	Deep Neural Networks	− 66.95
NTIS	− 2.63	TBL	− 0.05	Median-Constrained	− 0.25	Mean-models	1.24
INFL	− 2.58	IVSent3m	2.45	Mean-Constrained + IVSent	0.75	Median-models	2.12
SVAR	4.17	IVSent6m	2.45	Median-Constrained + IVSent	0.19		
		IVSent12m	1.59				

This table reports the results from the predictive regressions of individual factor models and of ensemble-factor models relative to the historical average naïve (benchmark) model. R^2_{OS} is the Campbell and Thompson (2008) out-of-sample R^2 statistic. If $R^2_{OS} > 0$, then mean squared prediction errors (MSPE) of $\bar{r}_{t+1}$, i.e. the predictive regression forecast, is smaller than for $\bar{r}_{t+1}$, i.e. the naïve forecast, indicating that the forecasting model outperforms the latter (benchmark) model. Panel A reports the results for the full out-of-sample period available (1965:1–2014:12) for all variables tested by Rapach *et al.* (2010). Panel B reports the results for the latest period within the entire out-of-sample history (2004:1–2014:12) and includes the three-month *IV-sentiment 90-110* factor (*IVSent*) in addition to the variables tested by Rapach *et al.* (2010).

regularization (shrinkage) applied as well as its adequacy to multicollinear systems. As the principal component regression also addresses multicollinearity problems and it performs quite poorly in the 1965:1–2014:12 sample, we conjecture that the main benefit delivered by the *Ridge regression* might be the shrinkage, which likely dampens the overfitting undergone by the 'kitchen sink' model. The *Random forest* model performs poorly, although, less bad than the 'kitchen sink' and the *Deep Neural Networks* models, suggesting that the structure imposed by constraint plus forecasting combination seems to add more value to predictions than being able to capture non-linear relationships. The *Deep Neural Networks* model performs as bad as the 'kitchen sink' model, likely due to overfitting. As we intentionally did not tune the *Random forest* and the *Deep Neural Networks* models much, the chance these models are overfitted is high. These two approaches are known by their potential for overfitting if regularization and stop-training procedures are not imposed. In summary, our results indicate the usefulness of some economic structure in modelling of the the equity risk premium and highlight the danger of overfitting when powerful machine learning methods are used without the proper handling.

Observing the evolution of $CSSED_{OS}$ for the median-based (restricted and unrestricted) ensemble models in Plot A of figure 4, we notice that both lines have slopes that are predominantly positive or flat. Positive slopes of the $CSSED_{OS}$ curve indicate that the ensemble model outperforms the benchmark out-of-sample. These $CSSED_{OS}$ lines match very closely the ones presented by Rapach *et al.* (2010) up to 2004, when their sample ends. The evolution of R^2_{OS} for our individual factors in figure 3 is also very similar to Rapach *et al.* (2010): some $CSSED_{OS}$ curves are positively sloped during certain periods, but often all factors display negatively sloped curves. The R^2_{OS} curves for the *IV Sent* factor is mostly positively sloped but relatively flat from 2004 to 2007, see the last plot in figure 4. These results reiterate the primary conclusion of Welch and Goyal (2008), Campbell and Thompson (2008), and Rapach *et al.* (2010): individual predictors that reliably outperform the historical average in forecasting the equity risk premium are rare but, once these models are sensibly restricted and aggregated in a multi-factor model, their out-of-sample predicting power improves considerably. This conclusion applies also to the inclusion of our *IV Sent* factor within the multi-factor model. Plot B of figure 4 shows that the $CSSED_{OS}$ curves for the model that includes the *IV Sent* factor are visibly steeper than the ones without it. Further, the findings in figure 4 indicate that restricted models are superior to unrestricted ones with either higher or less volatile $CSSED_{OS}$.

Even though the ensemble factor models do outperform the individual predictors, the red and black lines in Plots A and B of figure 5 are not always positively sloped, which is in line with Rapach *et al.* (2010). The R^2_{OS} curve is strongly positively sloped from 1965 to 1975, more moderately positively sloped from 1975 to 1992, negatively sloped from 1992 to 2000, and then slightly positive to flat until 2008, when it sharply drops amid the global financial crisis up to December 2014. The addition of our *IV Sent* factor in the ensemble

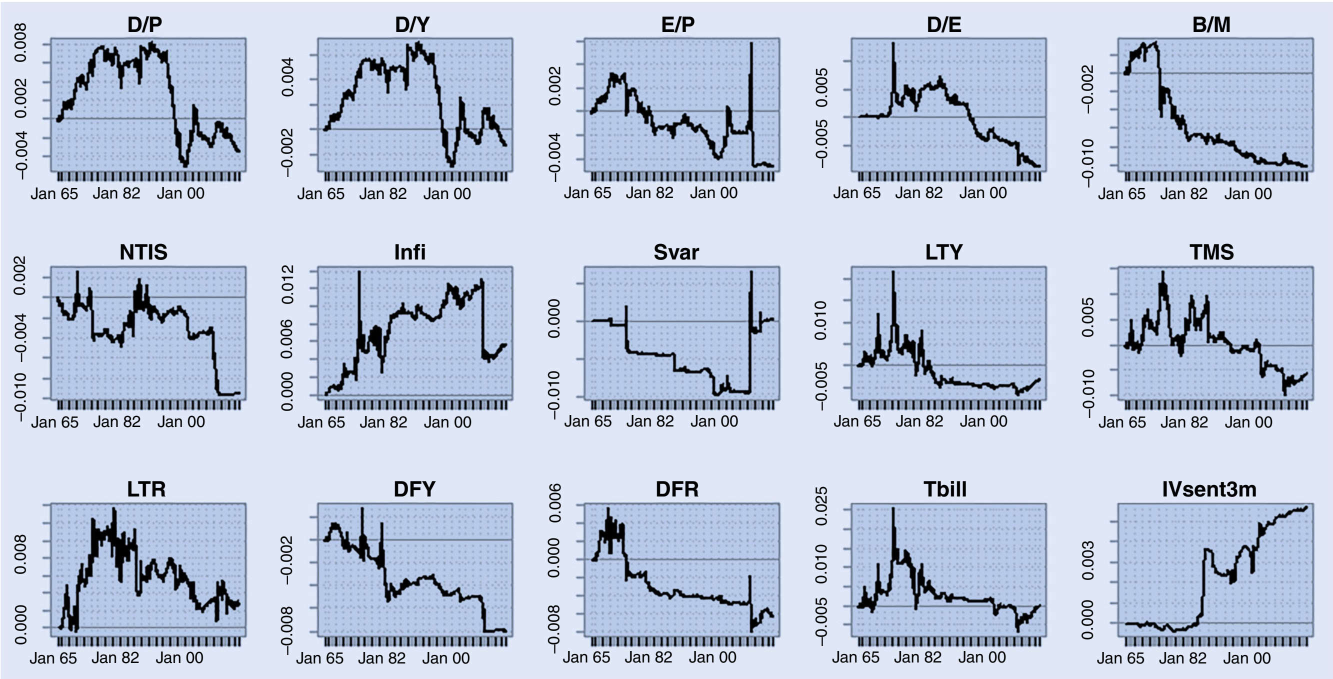

Figure 3. Cumulative Sum of Squared Error Differences of single factor predictive regressions. The lines in every plot depict the out-of-sample Cumulative Sum of Squared Errors Differences ($CSSED_{OS}$) calculated by equation (5) for the historical average benchmark-forecasting model minus the cumulative squared prediction errors for the single-factor forecasting models constructed by using 14 out of all the explanatory variables suggested by Welch and Goyal (2008), as well as the *IV-sentiment 90-110* factor with a three-month maturity. Positive values of $CSSED_{OS}$ mean that single-factor forecasting models that employ the Welch and Goyal (2008) factors and *IVsent* outperform the historical average benchmark-forecasting model.

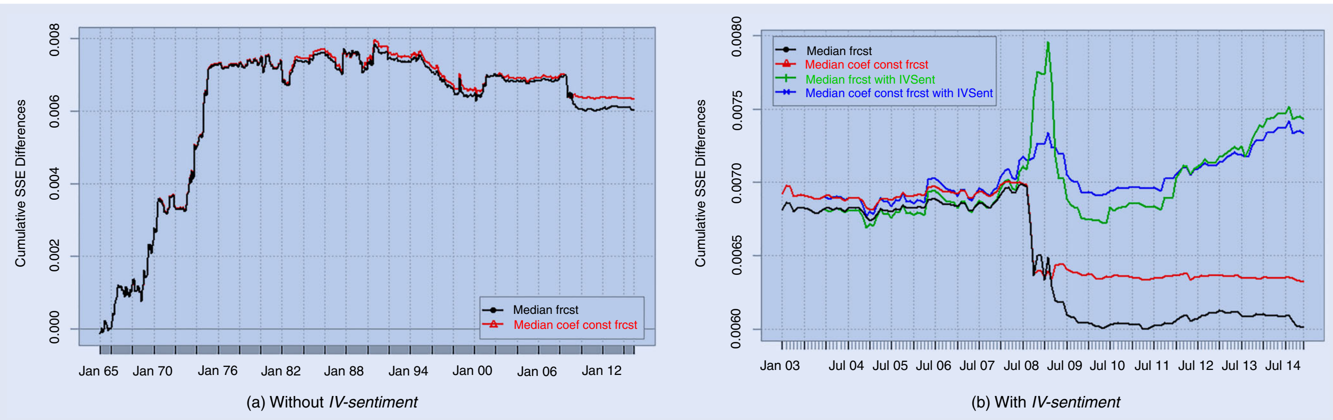

Figure 4. Cumulative Sum of Squared Error Differences of combined predictive regressions. The black line in Plot A depicts the Cumulative Sum of Squared Error Differences ($CSSED_{OS}$) for the historical average benchmark-forecasting model minus the cumulative squared prediction errors for the aggregated predictive regression-forecasting model construct by using 14 Welch and Goyal (2008) variables in univariate unrestricted models. The green and red lines in Plot A depict the same forecast evaluation statistic, i.e. the $CSSED_{OS}$, when such 14 univariate models are restricted as suggested by Campbell and Thompson (2008). The red line represents the $CSSED_{OS}$ when coefficients are constrained to have the same sign as the priors suggest. Plot B zooms in on the 2003:1–2014:12 period, where the black and red lines are the same as in Plot A, whereas the green and blue lines are the the $CSSED_{OS}$ when our *IV-sentiment* factor is added to the multifactor forecasts model for the unrestricted and restricted model, respectively. The forecasting period is 1965:1–2014:12 for all variables except *IVSent*, for which forecasts are only available from 2004:1–2014:12. Forecast aggregation in both models is done by calculating the mean of the $t + 1$ forecast from each individual predictive regression. (a) Without *IV-sentiment*. (b) With *IV-sentiment*.

model produces the blue and green lines in Plot B of figure 4. These new curves have an equally flat slope during the 2004 to 2008 period, while both experience a sharp rise since the beginning of 2008. These curves' profiles suggest that our *IV Sent* factor has considerably improved the out-of-sample performance of the ensemble model especially in times when the other factors broke down or did not provide an edge versus the historical average predictor. Thus, the inclusion of our *IV Sent* factor seems to revive the conclusion reached by the previous literature, where ensemble factor models are able to improve compared to individual factor models. At the same time, the recent poor performance of the ensemble models ex-*IV Sent* underscores that factor identification is still a major challenge for the specification of combined models. Overall, our empirical findings suggest that IV-based factors provide a relevant explanatory variable for the time-variation of equity returns.

3.3. IV-sentiment and equity factors

In this section we test whether the stream of returns produced by the *IV-sentiment* trading strategy is connected to (cross-sectional) equity factors. Our goal is to evaluate whether the *IV-sentiment* loads heavily on equity factors previously identified in the literature. Since the *IV-sentiment* aims to time entry and exit-points into the equity markets, it could potentially also be used by equity managers to time their beta exposure. Nevertheless, if this timing-strategy largely resembles equity factors, it should be less useful to equity portfolio managers.

We perform this analysis using equations (6a)–(6d), as well as univariate models using the individual factor employed in the following models:

$$IVSent_d = \alpha_d + (Mkt - RF)_d + SMB_d + HML_d + \epsilon_d, \quad (6a)$$

$$IVSent_d = \alpha_d + (Mkt - RF)_d + SMB_d + HML_d + WML_d + \epsilon_d, \quad (6b)$$

$$IVSent_d = \alpha_d + (Mkt - RF)_d + SMB_d + HML_d + WML_d + RMW_d + CMA_d + \epsilon_d, \quad (6c)$$

$$IVSent_m = \alpha_m + (Mkt - RF)_m + SMB_m + HML_m + WML_m + RMW_m + CMA_m + BAB_m + \epsilon_m, \quad (6d)$$

where the subscript $d = 1, 2, \ldots, D$ stands for daily returns, whereas the subscript $m = 1, 2, \ldots, M$ stands for monthly returns, both extending from January 2, 1999 to December 8, 2015. The first set of explanatory variables, used in equation (6a), are the market (*Mkt-Rf*), the size (*SMB*), and the value (*HML*) factors, as proposed by Fama and French (1992). Additionally, the profitability (*RMW*) and investment (*CMA*)† factor of Fama and French (2015), the momentum factor (*WML*) of Carhart (1997), and the low- versus high-beta (*BAB*), known as the 'Betting Against Beta' factor of Frazzini and Pedersen (2014) are used in equations (6b)–(6d).‡ The correlation structure of these factors estimated using our monthly data is reported in figure 5. In brief, it suggests that some cross-sectional equity factor can be highly positively or negatively correlated with each other but, more importantly, the *IV-sentiment* strategy seems only lowly correlated to all series.

Table 3 reports the results of equations (6a)–(6d). Note that the *IV-sentiment* has very little *Beta* exposure as the coefficients for the $(Mkt - RF)$ factor are close to zero across all models. This result matches our expectations as *IV-sentiment* has, in fact, a time-varying long or short exposure to the equity market. The *IV-sentiment* strategy also seems to have a large-cap tilt as the coefficient of *SMB* is often statistically significant and small or negative, ranging from -0.107 to 0.147. Again, this is an expected result as the *IV-sentiment* strategy is implemented in the US large cap universe, i.e. the S&P500 index. Coefficients for *HML* are also either low or negative, suggesting a growth tilt. *HML* is positive in the simpler models, i.e. in the univariate regression and in the Fama and French (1992) model, but negative in the more comprehensive models. This finding suggests the presence of multicollinearity in the model, which affects the estimated coefficient for *HML*. This effect is likely caused by the addition of the *RMW* factor, as it has a correlation of 0.5 in our sample (see figure 5).

Turning to the factors in equations (6b)–(6d) only, we find that *IV-sentiment* has negative exposure to the cross-sectional momentum factor (*WML*) consistently across all regressions. At first glance, this result makes sense as *IV-sentiment* is a mean-reversion strategy. Nevertheless, because the *IV-sentiment* reflects mean-reversion in the overall equity market, hence in a time-series fashion, rather than cross-sectionally, the expectation of a negative relation between these variables is ambiguous. Moskowitz *et al.* (2012) report that time-series momentum and cross-sectional momentum in the equity markets are strongly related though, which suggests that our original assumption that *IV-sentiment* is negatively correlated to *WML* holds. Among all factors, *WML* is almost the only one for which the statistical significance holds across all regressions. *WML* seems also to deliver, with around 2 percent, high explanatory power relative to the other factors used. This strong and robust negative link between *IV-sentiment* and *WML* reiterates our earlier suggestion that these two risk factors complement each other. And, by doing so, *IV-sentiment* might be able to mitigate some momentum crashes.

Moreover, the exposure of *IV-sentiment* to the profitability factor (*RMW*) is small and always negative, despite the fact that the coefficients are not statistically significant in the two multivariate models applied, only in the univariate regression. *IV-sentiment* is positively exposed to the investment factor (*CMA*) as its coefficients are significant across all regressions. We interpret that this positive relation with *IV-sentiment* relates to a higher frequency of reversals in periods when firm investments are low (likely during recessions or in the late economic cycle), which coincides with conservative firms outperforming aggressive ones. Besides, *IV-sentiment* loads negatively on the *BAB* factor, despite being only statistically significant in the univariate regression. This connection is argued to be linked to the profitability factor (*RMW*) by Fama and French (2016), which may help explain why both regressors are not statistically significant in the multivariate model, whereas they are strongly significant in the univariate regressions. In line with this suggestion, the estimated correlation between these two factors in our sample is 0.59 (see figure 5).

Last but not least, none of our regression models explains the variability *IV-sentiment* much as R^2 from equation (6d) is with 13 percent, at best, always low. This finding indicates that the *IV-sentiment* strategy is quite distinct from factors typically used by portfolio managers for single name equity management. Hence, as the *IV-sentiment* strategy embeds a timing approach for equity markets, which can be implemented via a dynamic exposure to market *Beta*, equity portfolio managers could enhance their strategies by making use of it.

3.4. Behavioral versus risk-sharing explanations

Another perspective of equity market dynamics provided by IV-based factors that are jointly extracted from single stock

† The Fama and French factors *SMB*, *HML*, *RMW* and *CMA* stand, respectively, for small minus big (size), high minus low (valuation), robust minus weak (profitability), and conservative minus aggressive (investments).

‡ The regressions that include the BAB factor have monthly frequency as this factor is not available in a daily frequency.

Table 3. Regression results: IV-sentiment and equity factors.

Panel A – Multivariate					Panel B – Univariate						
Intercept	0.000	0.000	0.000	0.007*	0.000	0.000	0.000	0.000	0.000	0.000	0.007
	(0.000)	(0.948)	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)
Mkt-RF	0.070***	0.042***	0.060***	0.072	0.073***						
	(0.010)	(0.011)	(0.012)	(0.104)	(0.010)						
SMB	0.134***	0.153***	0.136***	– 0.107		0.147***					
	(0.021)	(0.021)	(0.022)	(0.152)		(0.021)					
HML	0.080***	0.018	– 0.064***	– 0.271			0.086***				
	(0.019)	(0.021)	(0.024)	(0.180)			(0.020)				
WML		– 0.121***	– 0.141***	– 0.179*				– 0.134***			
		(0.015)	(0.015)	(0.094)				(0.013)			
RMW			– 0.042	– 0.130					– 0.137***		
			(0.029)	(0.220)					(0.024)		
CMA			0.244***	0.624**						0.107***	
			(0.036)	(0.245)						(0.029)	
BAB				– 0.186							– 0.215**
				(0.126)							(0.098)
R^2	2%	4%	5%	13%	1%	1%	0%	2%	1%	0%	M4%
F-stats	36.7	45.0	38.0	2.5	49.2	47.6	19.4	104.5	31.5	13.7	4.8
AIC	– 29771	– 29837	– 29879	– 430	– 29715	– 29713	– 29685	– 29769	– 29698	– 29680	– 429
BIC	– 29739	– 29798	– 29827	– 405	– 29696	– 29694	– 29666	– 29750	– 29678	– 29661	– 421

This table reports regression results for equations (6a), (6c) and (6d). The dependent variable is the stream of returns produced by the contrarian strategy based on our *IV-sentiment 90-110* indicator, while the explanatory variables are equity (cross-sectional) factors, namely: the market (Mkt-Rf), size (SMB), value (HML), profitability (RMW), investment (CMA), momentum (WML) and low- versus high-beta (BAB). Panel A reports the regression results in a multivariate setting, using three distinct model: (1) the Fama–French three-factor model, (2) the Fama–French three-factor model with the addition of the Carhart (1997) momentum factor, (3) the Fama–French five-factor model with the momentum factor and (4) the latter model with the addition of the BAB (Betting Against Beta) factor suggested by Frazzini and Pedersen (2014). Note that as the BAB factor is only available in monthly frequency, regression that contain such factor use monthly frequency, whereas data used in other regressions has daily frequency. We report standard errors in brackets. Asterisks ***, **, and * indicate significance at the one, five, and ten percent level, respectively.

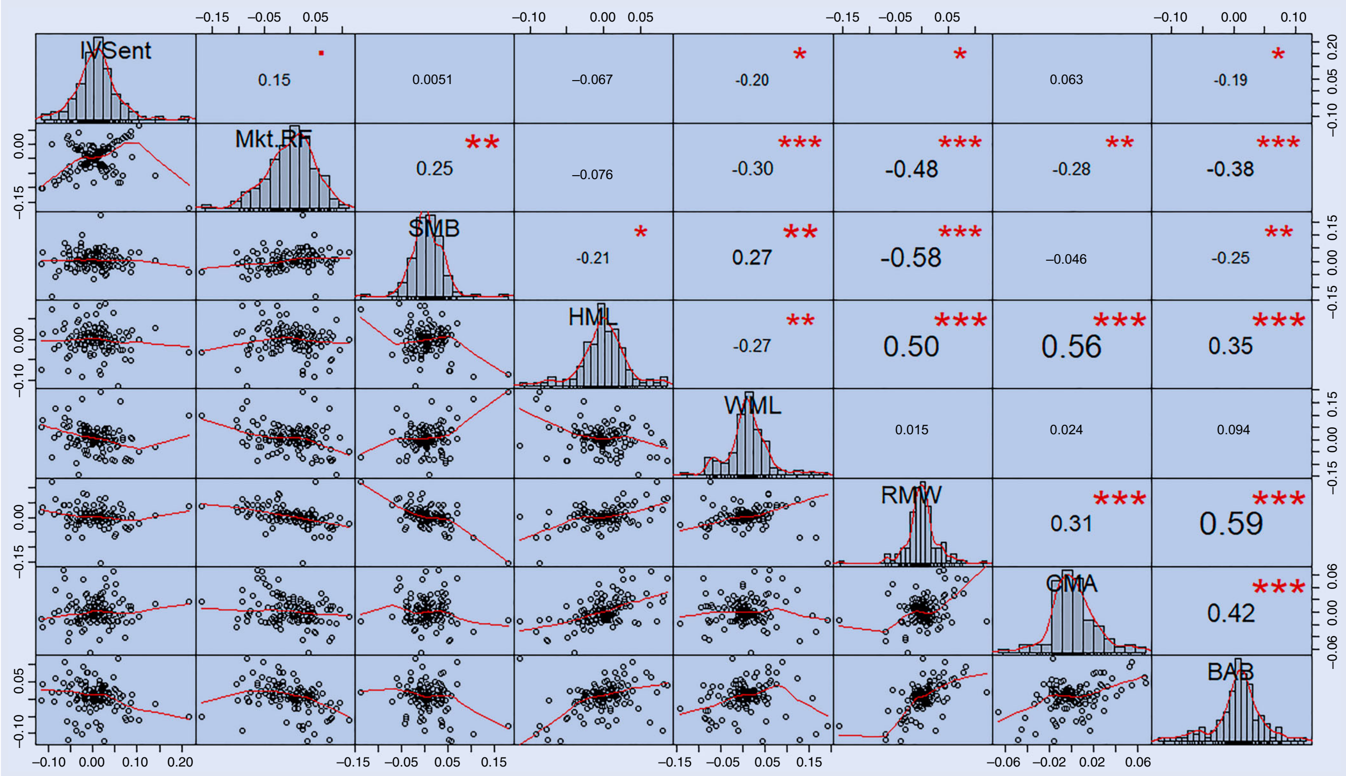

Figure 5. Correlation matrix between IV-sentiment factor and cross-sectional equity factors. The upper triangular part of the matrix above reports the correlation coefficient between pairs of cross-sectional equity factors and the *IV-sentiment* factor. These equity factors are the market (*Mkt-Rf*), the size (*SMB*) and the value (*HML*) factors, the profitability (*RMW*), the investment (*CMA*), the momentum factor (*WML*) and the 'Betting Against Beta' factor (*BAB*). The font size of coefficient reiterates its magnitude, whereas asterisks ***, **, and *indicate significance at the one, five, and ten percent level, respectively. In the diagonal, the histograms of factor returns are depicted. The lower triangular part of the matrix depicts scatter plots of the returns of the multiple pairs of factors.

and index options, is the implied correlation ($\bar{\rho}$). It is approximated by equation (7), which is derived in Appendix A.2:

$$\bar{\rho} \approx \frac{\sigma_I^2}{(\sum_{i=1}^n w_i \sigma_i)^2}, \tag{7}$$

where σ_I^2 is the variance of index options, σ_i is the volatility of $i = 1 \cdots n$ stocks in the index, and w_i is the stocks' weight in the index. The implied correlation measures the level of the average correlation between stocks that are constituents of an index. The IV of index options, i.e. (σ_I^2), can be matched by the one of single stock options, weighted by its constituents' loadings in the index, i.e. $(\sum_{i=1}^n w_i \sigma_i)^2$. Thus, if IV can be used as a measure of absolute expensiveness of an option, the implied correlation provides a relative valuation measure between the index and single stock options: a high (low) level of implied correlation means that index options are expensive (cheap) relative to single stock options.†

Table 4 Panel A presents descriptive statistics of the implied correlations between the index and single stock options' IV. The means and medians suggest that the implied correlation monotonically decreases with an increase in the moneyness level. The implied correlation means range from 0.30 to 0.65, which is wide given that these are averaged measures. Such a relative high dispersion of implied correlations is confirmed by their standard deviations, which are around 0.14. The most striking result is that the maximum implied correlation observed across all maturities and moneyness levels reaches 135 percent. Implied correlations above 100 percent are observed for many options, mostly for puts at the 80 and 90 percent moneyness levels. This finding implies that in order to match the weighted IV of puts on single stocks that are part of the S&P 500 index to the IV of a put on the index (with same levels of moneyness), an average correlation above 100 percent between the single stock put options is required. However, as correlation coefficients are bounded between −100 and +100 percent, these levels of implied correlation indicate irrational behavior by investors, who bid up index puts to levels that contradict market completeness.

We also find that trading in the opposite direction of such evident irrational investor behavior has been very profitable, as implied correlations higher than 100 percent were very effective as an entry point for contrarian strategies. Across the maturities and moneyness levels where we can observe such biased behavior, a sentiment strategy that buys the equity market when the implied correlation is above 100 percent and sells it when the implied correlation falls back to 50 percent, yields an average net IR of 0.35, with IRs ranging from 0.27 to 0.52.

The implied correlation means and medians provided by Panel A far exceed the same measures from realized average pair-correlations between the 50 largest constituents of

† Note that as $\sum_{i=1}^n w_i^2 \sigma_i^2$ is always positive, the approximation provided by equation (7) always overstates the true implied correlation. Given that $\sum_{i=1}^n w_i^2 \sigma_i^2$ is typically small, the current analysis remains valid. See Appendix A.2 for details.

Table 4. Implied and realized correlations.

Panel A – Implied correlations

Statistics \ Maturity, moneyness	3m 80%	3m 90%	3m ATM	3m 110%	3m 120%	6m 80%	6m 90%	12m 80%	12m 90%
Mean	0.65	0.56	0.45	0.35	0.3	0.64	0.56	0.6	0.54
Median	0.67	0.57	0.45	0.35	0.3	0.65	0.56	0.61	0.54
Minimum	0.24	0.18	0.12	0.07	0.03	0.26	0.21	0.26	0.22
Maximum	1.35	1.11	0.86	0.72	0.68	1.07	0.95	1.1	1.01
10th percentile	0.44	0.35	0.27	0.17	0.13	0.41	0.34	0.39	0.34
90th percentile	0.81	0.73	0.63	0.53	0.49	0.8	0.73	0.77	0.72
Standard deviation	0.15	0.14	0.14	0.13	0.14	0.14	0.14	0.14	0.13
Skew	− 0.46	− 0.39	0.1	0.29	0.29	− 0.6	− 0.38	− 0.26	− 0.2
Excess Kurtosis	0.6	− 0.02	− 0.37	− 0.38	− 0.66	0.09	− 0.18	0.03	− 0.22

Panel B – Realized correlations

Statistics \ Look-back period	30 Days	60 Days	90 Days	180 Days	720 Days
Mean	0.3	0.25	0.25	0.26	0.36
Median	0.27	0.22	0.24	0.25	0.31
Minimum	0	0.01	0	0.01	0.06
Maximum	0.84	0.69	0.67	0.61	0.74
10th percentile	0.1	0.05	0.04	0.07	0.08
90th percentile	0.54	0.47	0.48	0.52	0.71
Standard deviation	0.17	0.16	0.16	0.16	0.2
Skew	0.66	0.38	0.6	0.37	0.42
Excess Kurtosis	− 0.02	− 0.68	− 0.21	− 0.86	− 0.88

Panel A reports the descriptive statistics for the implied correlations between index options and single stock options for three month options at the 80, 90, ATM (100), 110, and 120 percent moneyness levels, and for six- and twelve-month options at the 80 and 90 percent moneyness levels over the full sample, which extends from January 2, 1998 to March 19, 2013. The implied correlation ($\bar{p}$) is approximated by the equation (7): $\bar{p} \approx \sigma_I^2/(\sum_{i=1}^n w_i\sigma_i)^2$, where σ_I^2 is the implied volatility of an index option and $\sum_{i=1}^n w_i\sigma_i$ is the weighted average single stock implied volatility, as in equation (A8l) of Appendix A. Panel B reports the descriptive statistics for the average pair-correlations for the 50 largest constituents of the S&P500 index calculated over the same sample, which extends from January 2, 1998 to March 19, 2013.

the S&P 500 index as of February 14, 2014, as provided in Panel B. Such average pair-correlations range from 0.25 to 0.36 when look-back periods of 30, 60, 90, 180, and 720 days are evaluated, which is substantially lower than most average implied correlations posted for the different option maturity and moneyness levels reported in Panel A. In fact, the average realized correlations are often below the 10th percentile of the implied correlation for some options' maturity and moneyness levels. The 90th percentile of realized correlations often match the average implied correlations reported. The maximum realized correlations are at most 84 percent, using an extremely short look-back of 30 days, much lower than the 135 percent observed for implied correlations. These empirical findings strongly suggest that implied correlations substantially overshoot realized ones. Similarly, the implied correlation reaches sometimes values as low as three percent for some options, especially on the call side (above ATM moneyness). This finding is also low when compared to put options. The minimum historical correlations from OTM puts is 0.18, whereas for call options it is 0.03. The fact that those extremely low values of the implied correlation from call options largely undershoots implied correlations from puts may also suggest less than fully rational pricing on the call side. It indicates that single stock options are expensive relative to index calls, which matches our postulation that individual investors use single stock calls to speculate on the upside.

Despite the strong evidence of irrational behavioral by investors provided by the extreme levels of implied correlation, which indirectly links to the IV skew being at extreme levels at times, we conjecture that this strategy delivers long-run positive returns may also have a risk-bearing explanation. Reversal strategies such as the ones designed by us earn attractive long-term risk-adjusted returns, but are highly dependent on equity markets at the tail (see table 1, Panel C). Additionally, *IV-sentiment*-based reversal strategies experience the largest daily drawdowns among all strategies evaluated (see table 1, Panel A). Thus, their attractive risk-adjusted returns are, partially, compensation for downside risk. Therefore, the risk borne by investors that bet on reversals in equity markets is the risk of poor timing of losses (Harvey and Siddique 2000) and downside risk (Ang *et al.* 2006). In brief, betting on equity market reversals is a risky activity.

We note that this rational explanation for excesses in sentiment is also linked to the *limits-to-arbitrage* literature: as investors have finite access to capital (Brunnermeier and Pedersen 2009) and feedback trading can keep markets irrational for a long period of time (De Long *et al.* 1990), contrarian strategies aiming to exploit the effect of irrational trading are not without risk. For example, once bearish sentiment seems excessive, the risk of betting on a reversal may be tolerable only to a few investors, because (1) higher volatility drags investors' risk budget usage closer to its limits, and (2) access to funding is limited. The ability to 'catch a knife falling' in the equity markets is not suitable for all investors, as it involves high risk. Contrarian strategies are, then, mainly accessible to investors that have enough capital or funding liquidity. Similar considerations are career risk (Chan *et al.* 2002), negative skewness of returns (Harvey and Siddique 2000), poor timing of losses (Campbell and Cochrane 1999, Harvey and Siddique 2000), and risk aversion of market makers (Garleanu *et al.* 2009). One final element in the characterization of reversals as a compensation for risk is the presence of correlation risk priced in index options (see Krishnam 2009, Driessen *et al.* 2013), which is present in assets that perform well when market-wide correlations are higher than expected.

4. *IV-sentiment* and overweight of tails

Overweight of tail events is claimed to largely impact the pricing of OTM index puts and OTM single stock calls (Barberis and Huang 2008, Polkovnichenko and Zhao 2013). Papers that empirically test this hypothesis are very few, such as Polkovnichenko and Zhao (2013), Dierkes (2009), and Félix *et al.* (2019). Given that our proposed sentiment measure is manufactured from OTM IVs, in this section we aim to identify (1) whether *IV-sentiment* is closely linked to formal metrics of overweight of tail events estimated from options as well as (2) whether such measure is linked to sentiment and (3) forward returns.

4.1. *Estimating overweight of tails*

Overweight of small probabilities is embedded in the cumulative prospect theory (CPT) model by means of the weighting function of the probability of prospects, using parameters δ and γ for the left (losses) and right (gains) side of the return distribution, respectively. δ and $\gamma < 1$ imply overweight of small probabilities, whereas δ and $\gamma > 1$ imply underweight of small probabilities, and δ and γ equal to 1 means neutral weighting of prospects (Tversky and Kahneman 1992).

Our methodology builds on the assumption that investors' subjective density estimates should correspond, on average,† to the distribution of realizations (Bliss and Panigirtzoglou 2004). Thus, estimating CPT probability weighting function parameters δ and γ is only feasible if two basic inputs are available: the CPT subjective density function and the distribution of realizations, i.e. the empirical density function (EDF). The methodology applied by us to estimate these two parameters comprises of: (1) estimating the returns' risk-neutral density from option prices using a modified (Figlewski 2010) method; (2) estimating the partial CPT density function using the CPT marginal utility function; (3) 'undoing' the effect of the probability weighting function (w) to obtain the CPT subjective density function; (4) simulating time-varying empirical return distributions using the Rosenberg and Engle (2002) approach; and (5) minimizing the squared difference of the tail probabilities of the CPT and EDF to obtain daily optimal δ's and γ's.

Our starting point for obtaining the CPT probability weighting function parameters δ and γ is the estimation of RND from IV data, by first applying the Black-Scholes model to our IV data to obtain options prices (C) for the S&P 500 index. Once our data is normalized, so strikes are expressed in terms of percentage moneyness, the instantaneous price level of the S&P 500 index (S_0) equals 100 for every period for which we would like to obtain implied returns. Contemporaneous dividend yields for the S&P 500 index are used for the calculation of P as well as the risk-free rate from three-, six- and twelve-month T-bills. Because we have IV data for five levels of moneyness, we implement a modified Figlewski (2010) method for extracting the RND structure. The main advantage of the Figlewski (2010) method over other techniques is that it extracts the body and tails of the distribution separately, thereby allowing for fat tails.

Once the RND is estimated, we translate it into the subjective density function, a *real-world* probability distribution, via the pricing kernel as follows:

$$\frac{f_Q(S_T)}{f_P(S_T)} = \Lambda \frac{U'(S_T)}{U'(S_t)} \equiv \varsigma(S_T). \tag{8}$$

We further manipulate equation (8) so to directly relate the original EDF to the CPT subjective density function, by 'undoing' the effect of the CPT probability distortion functions within the PCPT density function. The relation between EDF and the CPT density function is given by equation (9); its derivation from equation (8) is provide in Appendix 1:

$$\underbrace{f_P(S_T)}_{\text{EDF}} = \underbrace{\frac{\frac{f_Q(S_T)}{v'(S_T)}}{\int \frac{f_Q(x)}{v'(x)}dx}(w^{-1})'(F_P(S_T))}_{\text{CPT density function}} \tag{9}$$

Thus, once the relation between the probability weighting function of the EDF and the PCPT density is established, as in equations (A5) and (A6) in the Appendix, we eliminate the weighting scheme affecting returns by applying the inverse of such weightings to the subjective density function without endangering such equalities, as in equation (9).

As the RND is converted into the subjective density function, we must also estimate daily empirical density functions (EDF). We built time-varying EDFs from an invariant component, the *standardized innovation density*, and a time-varying part, the lagged conditional variance ($\sigma^2_{t|t-1}$) produced by an EGARCH model. We first define the *standardized innovation*, being the ratio of empirical returns and their conditional standard deviation ($\ln(S_t/S_{t-1})/\sigma_{t|t-1}$) produced by the EGARCH model. From the set of standardized innovations produced, we can then estimate a density shape, i.e. the *standardized innovation density*. The advantage of such a density shape versus a parametric one is that it may include the typically observed fat tails and negative skewness, which are not incorporated in simple parametric models, e.g. the normal distribution. This density shape is invariant and it is turned time-varying by multiplication of each standardized innovation by the EGARCH conditional standard deviation at time t,

† This assumption implies that investors are somewhat rational, which is not inconsistent with the CPT-assumption that the representative agent is less than fully rational. The CPT suggests that investors are biased, not that decision makers are utterly irrational to the point that their subjective density forecast should not correspond, on average, to the realized return distribution.

which is specified as follows:

$$\ln(S_t/S_{t-1}) = \mu + \epsilon_t, \epsilon \sim f(0, \sigma^2_{t\,|\,t-1}) \tag{10a}$$

and

$$\sigma^2_{t\,|\,t-1} = \omega_1 + \alpha\epsilon^2_{t-1} + \beta\sigma^2_{t-1\,|\,t-2} + \vartheta\,\mathrm{Max}[0, -\epsilon_{t-1}]^2, \tag{10b}$$

where α captures the sensitivity of the conditional variance to lagged squared innovations (${\epsilon^2}_{t-1}$), β captures the sensitivity of the conditional variance to the conditional variance (${\sigma^2}_{t-1\,|\,t-2}$), and ϑ allows for the asymmetric impact of lagged returns ($\vartheta\mathrm{Max}[0, -\epsilon_{t-1}]^2$). The model is estimated using maximum log-likelihood where innovations are assumed to be normally distributed.

Up to now, we produced a one-day horizon EDF for every day in our sample but we still lack time-varying EDFs for the three-, six-, and twelve-month horizons. Thus, we use bootstrapping to draw 1,000 paths towards these desired horizons by randomly selecting single innovations (ϵ_{t+1}) from the one-day horizon EDFs available for each day in our sample. We note that once the first return is drawn, the conditional variance is updated ($\sigma^2_{t-1\,|\,t-2}$) affecting the subsequent innovation drawings of a path. This sequential exercise continues through time until the desired horizon is reached. To account for drift in the simulated paths, we add the daily drift estimated from the long-term EDF to drawn innovations, so that the one-period simulated returns equal $\epsilon_{t+1} + \mu$. The density functions produced by the collection of returns implied by the terminal values of every path and their starting points are our three-, six-, and twelve-month EDFs. These simulated paths contain, respectively, 63, 126, and 252 daily returns. We note that by drawing returns from stylized distributions with fat-tails and excess skewness, our EDFs for the three relevant horizons also embed such features. This estimation method for time-varying EDF is based on Rosenberg and Engle (2002).

Finally, once these three time-varying EDFs are estimated for all days in our sample, we estimate δ and γ for each of these days using equations (11) and (12):

$$w^+(\gamma, \delta = \gamma) = \mathrm{Min}\sum_{b=1}^{B} W_b(EDF^b_{prob} - CPT^b_{prob})^2, \tag{11}$$

$$w^-(\delta, \delta = \gamma) = \mathrm{Min}\sum_{b=1}^{B} W_b(EDF^b_{prob} - CPT^b_{prob})^2, \tag{12}$$

where EDF^b_{prob} and CPT^b_{prob} are the probability contained in each bin b that divides the range of the empirical and CPT probability distributions. W_b are weights given by $1/(1/\sqrt{2\pi})\int_{0.5}^{\infty} e^{-x^2/2}\,\mathrm{d}x = 1$, the reciprocal of the normalized normal density function (above its median), split in the same total number of bins (B) used for the EDF and CPT. Parameters δ and γ are constrained by an upper bound of 1.75 and a lower bound of -0.25. The weights applied in these optimizations are due to the higher importance of matching probability tails in our analysis than to the body of the distributions.

4.2. Time-varying overweight of tails

In the following, we evaluate the dynamics of the overweighting of tails within the single stock and index option markets. Descriptive statistics of the CPT's estimated δ and γ parameters via the methodology presented in Section 2 are provided in table 5.

We report summary statistics of the estimated γ for three-, six- and twelve-month options in Panel A for the right tail from single stock options. The median and mean time-varying γ estimates for three-month options are 0.89 and 0.91, respectively, which considerably exceed the parameter value of 0.61 in Tversky and Kahneman (1992). This finding suggests that overweight of small probabilities is present within the pricing of short-term single stock call options, but to a much lesser extent than provided by the theory. The results in Panel A also show that γ is highly time-varying and strongly sample dependent. Overweight of small probabilities in the single stock option market is very pronounced from 1998 to 2003 (present at 97 percent of all times), but infrequent from 2003 to 2008 (present at only 35 percent of all times). Our γ-estimates from three-month options range from 0 to 1.75 and the standard deviation is 0.23. In Panel B, we report summary statistics of the estimated δ from index options for the left tail. For δ estimated from three-month options, the median and mean estimates are both 0.68, implying a probability weighting that roughly matches the one in the CPT, which calibrates δ at 0.69. The δ-estimates are also time-varying, however, their standard deviation (0.08) is more than three times lower than for the γ-estimates. The range of δ-estimates is also much narrower than for γ, as it is between 0.29 and 1.01. In contrast to the γ-estimates, our δ-estimates reflect a consistent overweight of small probabilities across all sub-samples.

At the six-month maturity, overweight of small probabilities for γ seems even less acute than suggested by theory and by the three-month options findings. The median and mean γ estimates for this maturity are 0.99 and 0.96, respectively. The distribution of γ is somewhat skewed to the right, i.e. towards a less pronounced overweight of small probabilities, as the median is higher than the mean. The 75th quantile of γ (1.14) suggests an underweighting of probabilities already. For index options with six-month maturity, the estimated δ indicates an even more pronounced overweight of small probabilities (both the mean and median δ equal 0.60) than for three-month options. Overweight of small probabilities is again documented across all samples for δ but not for γ, in which overweight of small probabilities is more frequent than underweight of small probabilities only in the 1998–2003 sample.

The γ estimates for the twelve-month maturity tend even more towards probability underweighting than the six-month ones. The median γ is 1.03, whereas the mean γ is 1.01. Overweight of small probabilities appears in only 41 percent of all times in the overall sample and is roughly non-existent in the 2003–2008 sample. Differently, the mean and median for the δ estimates from index options are 0.47 and 0.40, respectively, indicating an even stronger overweight of small probabilities than for single stock options and other maturities. We argue that such a pattern could be caused by institutional investors

Table 5. Descriptive statistics.

Panel A – Gamma												
Maturity	Min	25% Q	Median	Mean	75% Q	Max	StDev	% $\gamma < 1$	% $\gamma < 1$ (98-03)	% $\gamma < 1$ (03-08)	% $\gamma < 1$ (08-13)	RSS
3 months	–	0.74	0.91	0.89	1.04	1.75	0.23	64%	97%	35%	59%	0.0209
6 months	–	0.81	0.99	0.96	1.14	1.75	0.28	52%	92%	18%	46%	0.0170
12 months	0.04	0.91	1.03	1.01	1.14	1.75	0.22	41%	83%	11%	29%	0.0225
Panel B – Delta												
Maturity	Min	25% Q	Median	Mean	75% Q	Max	StDev	% $\delta < 1$	% $\delta < 1$ (98-03)	% $\delta < 1$ (03-08)	% $\delta < 1$ (08-13)	RSS
3 months	0.29	0.64	0.68	0.68	0.72	1.01	0.08	100%	100%	100%	100%	0.0579
6 months	0.30	0.54	0.60	0.60	0.65	1.75	0.10	100%	100%	100%	100%	0.0198
12 months	–	0.40	0.45	0.47	0.52	1.75	0.10	100%	100%	100%	100%	0.0169

This table reports the summary statistics of the estimated cumulative prospect theory (CPT) parameters gamma (γ) from the single stock options market and delta (δ) from the index option market for each day in our sample as well as the optimizations' residual sum of squares (RSS). The parameters γ and δ define the curvature of the weighting function for gains and losses, respectively, which leads the probability distortion functions to have inverse S-shapes. The γ and δ parameters close to unity lead to weighting functions that are close to unweighted (neutral) probabilities, whereas parameters close to zero indicates large overweight of small probabilities. Panel A reports the summary statistics of gamma (γ) when we assume a parameter of risk aversion (λ) equal to 2.25 (the standard CPT parametrization). Panel B reports the summary statistics of delta (δ) under the same risk aversion assumption. Column headings % $\gamma < 1$ and % $\delta < 1$ report the percentage of observations in which parameters γ and δ are smaller than one, i.e. the proportion of the sample in which overweight of small probabilities is observed. We report this metric for the full sample as well as for three equal-sized splits of our full samples, namely: 98-03, from 1998-01-05 to 2003-01-30; 03-08, from 2003-01-31 to 2008-02-21; and 08-13, from 2008-02-22 to 2013-03-19.

buying long-term protection, as twelve-month OTM index options are less liquid than short-term ones.

OTM index puts seem to be structurally expensive from the perspective of overweight of small probabilities, even though the degree of overvaluation varies in time. Concurrently, OTM single stock options are only occasionally expensive, and clustered in specific parts of our sample, e.g. during the 1998–2003 period. Our results fit nicely within the seminal literature, for instance with Dierkes (2009), Kliger and Levy (2009), and Polkovnichenko and Zhao (2013), regarding the index option market, and with Félix *et al.* (2019) regarding the single stock option market.

4.3. *Overweight of tails and sentiment*

In order to evaluate how time-variation in overweight of small probabilities relates to sentiment, we run regressions between our proxies for overweight of tails, the Baker and Wurgler (2007) sentiment measure, and other explanatory control variables. Since we aim to combine overweight of small probabilities parameters from both index options (bearish sentiment) and single stock options (bullish sentiment), we use *Delta minus Gamma spread*, δ - γ, as the explained variable. *Delta minus Gamma spread* captures the overweighting of small probabilities from both index options and single stock, because δ is the CPT tail overweight parameter estimated from the single stock market, and γ is the equivalent parameter estimated from the index option market. The explanatory variables in these regressions are (1) the Baker and Wurgler (2007) sentiment measure,† (2) the percentage of bullish investors minus the percentage of bearish investors given by the survey of the American Association of Individual Investors (AAII), (3) a proxy for individual investors' sentiment (see Han 2008), and (4) a set of control variables among the ones tested by Welch and Goyal (2008)‡ as potential forecasters of the equity market. The data frequency used is monthly, as this is the highest frequency in which the Baker and Wurgler (2007) sentiment factor and the Welch and Goyal (2008) data set are available. Our sample starts in January 1998 and ends in February 2013.§ The OLS regression model is:

$$\begin{aligned} DGspread[\tau]_t = c + SENT_t + IISENT_t + E12_t + B/M_t \\ + NTIS_t + TBL_t + INFL_t + CORPR_t \\ + SVAR_t + CSP_t + \epsilon_t, \qquad (13) \end{aligned}$$

where τ is the option horizon, *DGspread* is the *Delta minus Gamma spread*, *SENT* is the Baker and Wurgler (2007) sentiment measure, *IISENT* is the AAII individual investor sentiment measure, $E12$ is the twelve-month moving sum of earnings of the S&P 5000 index, *B*/*M* is the book-to-market ratio, *NTIS* is the net equity expansion, *TBL* is the risk-free rate, *INFL* is the annual INFLation rate, *CORPR* is the corporate spread, *SV AR* is the stock variance, and *CSP* is the cross-sectional premium. We also run the following univariate models for each explanatory factor separately to understand

† Available at http://people.stern.nyu.edu/jwurgler/.

‡ The complete set of variables provided by Welch and Goyal (2008) that is employed here is discussed in Appendix B. To avoid multicollinearity in our regression, we exclude all variables that correlate more than 40 percent with others.

§ This sample is only possible because Welch and Goyal (2008) and Baker and Wurgler (2007) have updated and made available their datasets after publication.

their individual relation with the *Delta minus Gamma spread*:

$$DGspread[\tau]_t = \alpha_i + \beta_i x_{i,t} + \epsilon_t, \quad (14)$$

where x represents the 10 explanatory variables specified in equation (13), thus $i = 1 \cdots 10$.

Table 6 Panel A reports the results of equation (13), estimated across three maturities for *Delta minus Gamma spread*. The explanatory power of the multivariate regression is very high, ranging from 36 to 57 percent. As expected, *SENT* is positively linked to *Delta minus Gamma spread* and statistically significant across the three- and six-month maturities. This suggests that high sentiment exacerbates overweight of small probabilities measured as *Delta minus Gamma spread*. However, this relation is negative and not significant at the twelve-month maturity. The univariate regressions of *SENT* confirm the positive link between sentiment and *Delta minus Gamma spread* at shorter maturities. Once again, this relation is not present at the twelve-month horizon. The explanatory power of *SENT* in the univariate setting is also high for the three- and six-month horizons, with 17 and 32 percent, respectively. This result strengthens our hypothesis that overweight of small probabilities increases at higher levels of sentiment and that sentiment seems to have a strong link to probability weighting by investors as priced by index puts and single stock call options. This finding, however, applies to the three- and six-month horizons only since the twelve-month univariate regression has a R^2 of zero.†

IISENT is also positively connected to *Delta minus Gamma spread* in the multivariate regression at the three- and six-month horizons but negatively at the twelve-month horizon. These results are confirmed by the univariate regressions, as *IISent* is positively linked to *Delta minus Gamma spread* at the three- and six-month horizons. Explanatory power of these regressions is with 6 percent for both the three- and six-month maturities, relatively high. For the twelve-month maturity in the univariate regression, *IISENT* is negatively linked to *Delta minus Gamma spread* and is statistically significant.

Once we analyze the other control variables in our regression, we observe that the results are less stable than for the sentiment proxies. Table 6 indicates that some signs of control variables change in both the multivariate and univariate regressions. *TBL* is the only control variable that remains statistically significant and keeps its sign across the multivariate and univariate models. The explanatory power of *TBL* is 21 percent in the univariate setting, whereas the other independent variable with high explanatory power is book-to-market with 27 percent. *B/M* is only statistically significant in the three-month maturity of the multivariate regressions. *NTIS* is negatively and significantly linked to *Delta minus Gamma spread* in the univariate setting as well as in the multivariate regression in the twelve-month maturity. *SVAR* is negatively and significantly linked to *Delta minus Gamma spread* in the univariate regression but in the multivariate regression this result is not observed. Overall, these empirical findings suggest that fundamentals have a relatively unstable link to the *Delta minus Gamma spread*.

We note that the high stability of the relation between the sentiment factors and the *Delta minus Gamma spread* within the multivariate regressions provides evidence that sentiment and overweight of small probabilities are strongly connected.

4.4. Relating overweight of tails to IV-sentiment, IV skews and higher moments of the RND

In a next step, we assess the relationship between IV-sentiment, *Delta minus Gamma spread*, and higher moments (skewness and kurtosis) of the RND implied by options and IV skew measures. We undertake this analysis to understand to which extent *Delta minus Gamma spread* is connected to *IV-sentiment* and other metrics seemingly derived from IV.

We expect the existence of a positive link between the estimated *Delta minus Gamma spread* and IV skew measures, because the presence of fat tails in the RND is a pre-condition for overweight of tail probabilities and a corollary of OTM's IVs to be rich versus at-the-money (ATM) IVs. Similarly, we observe negative skewness and fat-tails in RNDs only if OTM options are expensive versus ATM options and vice-versa.‡ Consequently, γ and δ are likely to be smaller than one (overweight of small probabilities), and *Delta minus Gamma spread* differs from zero if OTM options are expensive versus ATM options, which supports the use of IV skew as another proxy for overweight of tails.

Beyond IV-sentiment, the IV skew measures used are the standard measures: (1) IV 90 percent (moneyness) minus ATM, (2) IV 80 percent minus ATM from index options (which captures bearish sentiment), (3) IV 110 percent minus ATM, and (4) IV 120 percent minus ATM from single stock calls (which captures bullish sentiment).

We assess the isolated relationship between *Delta minus Gamma spread* and *IV-sentiment*, higher moments of the RND, and (standard) IV skews measures using the univariate models presented by equations (15)–(18). These models are estimated with OLS, where Newey-West standard errors are used for statistical inference. Our daily regression samples start on January 2, 1998 and end on March 19, 2013.

$$DGspread[\tau] = \alpha_t\left[\frac{K}{S}\right] + IVSent_t\left[\frac{K}{S};\tau\right] + \epsilon_t, \quad (15)$$

$$DGspread[\tau] = \alpha_t + KURT_t^m(\tau) + \epsilon_t, \quad (16)$$

$$DGspread[\tau] = \alpha_t + SKEW_t^m(\tau) + \epsilon_t, \quad (17)$$

$$DGspread[\tau] = \alpha_t\left[\frac{K}{S}\right] + IVSKEW_t\left[\frac{K}{S};\tau\right] + \epsilon_t, \quad (18)$$

† Our results suggest that overweight of small probabilities is much less pronounced at the twelve-month options, at least for single stock options (see table 5), and that the *Delta minus Gamma spread* is disconnected to sentiment (*SENT*) at this same maturity. Beyond that, *IV-sentiment* using twelve-month options is not reliable as an active management signal (see figure 1), likely by not properly capturing sentiment but rather reflecting risk-neutral pricing. That said, we believe our twelve-month metrics might still be very useful from a risk management perspective, as the consideration of risks in the long-run is also important.

‡ While these relations are widely acknowledged, Longstaff (1995) provide a formal theorem for the link between IV skew and risk-neutral moments, whereas Bakshi *et al.* (2003) offer a comprehensive empirical test of this proposition for index options.

Table 6. Regression results: *Delta minus Gamma spread*.

Panel A - Multivariate				Panel B - Univariate													
Maturity	3m	6m	12m	3m	6m	12m	3m	6m	12m	6m	6m	6m	6m	6m	6m	6m	6m
Intercept	0.003	– 0.491***	– 0.490***	– 0.063***	– 0.369***	– 0.520***	– 0.064***	– 0.365***	– 0.508***	– 0.048	0.131***	– 0.055***	– 0.121***	– 0.055***	– 0.053***	– 0.039***	– 0.052***
	(0.056)	(0.037)	(0.058)	(0.010)	(0.008)	(0.013)	(0.011)	(0.010)	(0.012)	(0.031)	(0.031)	(0.011)	(0.015)	(0.013)	(0.011)	(0.012)	(0.011)
SENT	0.030*	0.064***	– 0.024	0.071***	0.097***	– 0.003											
	(0.017)	(0.013)	(0.019)	(0.014)	(0.016)	(0.016)											
IISENT	0.041	0.096**	– 0.106**				0.123***	0.125**	– 0.124***								
	(0.047)	(0.038)	(0.048)				(0.044)	(0.050)	(0.043)								
E12	0.000	– 0.003	– 0.028***							– 0.001							
	(0.006)	(0.004)	(0.007)							(0.006)							
B/M	– 0.364*	0.125	0.163								– 0.737***						
	(0.217)	(0.132)	(0.211)								(0.130)						
NTIS	0.560	0.259	– 0.814									1.075**					
	(0.391)	(0.285)	(0.523)									(0.440)					
TBL	0.013	0.036***	0.029***										0.030***				
	(0.008)	(0.006)	(0.009)										(0.006)				
INFL	0.453	1.843	2.311											1.784			
	(2.507)	(1.885)	(2.176)											(3.350)			
CORPR	0.225	0.233	0.044												0.128		
	(0.285)	(0.202)	(0.273)												(0.472)		
SVAR	– 1.426	3.519***	3.470*													– 3.376**	
	(1.331)	(1.153)	(1.982)													(1.307)	
CSP	– 0.125	0.198	0.261														0.029
	(0.136)	(0.122)	(0.235)														(0.197)
R^2	36%	57%	30%	17%	32%	0%	6%	6%	5%	0%	27%	4%	21%	0%	0%	4%	0%
F-stats	8.2	19.5	6.4	32.5	72.9	0.0	9.1	9.4	8.0	0.0	58.0	7.1	40.7	0.7	0.2	6.1	0.0
AIC	– 308.1	– 369.1	– 273.2	– 326.1	– 186.0	34.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
BIC	– 274.4	– 335.4	– 239.6	– 320.0	– 179.9	40.3	0.0	0.0	0.0	0.0	0.1	0.4	0.0	2.2	0.3	1.4	0.2

Panel A reports the regression results for equation (13) in a multivariate setting. The dependent variable is *Delta minus Gamma spread* (δ - γ),while as explanatory variables we specify: (1) the Baker and Wurgler (2007) sentiment measure (*SENT*), (2) the individual investor sentiment (*IISENT*), and (3) the explanatory variables used by Welch and Goyal (2008), while excluding factors that correlate to each other in excess of 40 percent (see Appendix 2 for the full list of variables). Panel B reports the regression results for (14), in an univariate setting, in which *Delta minus Gamma spread* is regressed on the same set of explanatory variables. We report Newey-West adjusted standard errors in brackets. Asterisks ***, **, and * indicate significance at the one, five, and ten percent level, respectively.

where K/S is the moneyness level of the option, τ is the option horizon, *DGspread* is the *Delta minus Gamma spread*, *IV Sent* is our *IV-sentiment* measure, *SKEW* is the RND return skewness implied by options, *KURT* is the RND return kurtosis implied by options, and *IV SKEW* is the single market IV skew measure, for both index option and single stock option markets. We note that the superscript m for the variables *KURT* and *SKEW* distinguishes RND kurtosis and skewness obtained from either index options ($m = io$) or single stock options ($m = sso$).

We estimate multivariate models of *Delta minus Gamma spread* regressed on RND skewness, kurtosis, IV skews and *IV-sentiment* to better understand the relation between these measures jointly and overweight of small probabilities:

$$DGspread[\tau] = \alpha_t \left[\frac{K}{S}\right] + SKEW_t^m(\tau) + KURT_t^m(\tau) + IVSent_t\left[\frac{K}{S};\tau\right] + \epsilon_t. \quad (19)$$

Table 7 Panel A reports the estimates of equations (15)–(18), when *DGspread* is regressed on RND moments, IV skews, and *IV-sentiment 90-110* in a univariate setting. The empirical findings indicate that *IV-sentiment* is the variable that explains *DGspread* the most across all maturities. The explanatory power of *IV-sentiment* is not only the highest but it is also the most consistent factor, as its R^2 ranges from 30 to 46 percent. *IV-sentiment* is negatively related to *DGspread*. Such a negative sign of the *IV-sentiment* regressor was expected because *DGspread* rises with higher bullish sentiment, whereas higher *IV-sentiment* suggests a more pronounced bearish sentiment. Risk-neutral skewness and kurtosis also strongly explains *DGspread* (by roughly 30 percent), though only within the three-month maturity. Skewness and kurtosis explain *DGspread* by roughly 10 percent for six-month options, and 7 percent for twelve-month ones. The coefficient signs are in line with our expectations since high levels of RND skewness are associated with high *DGspread* (a bullish sentiment signal), while low levels of RND kurtosis (less pronounced fat-tails) are associated with high *DGspread*.† In contrast, standard IV skews explain with only between 0 and 4 percent very little of *DGspread* within the three-month maturity. At longer maturities, the IV skews are able to better explain *DGspread*, however, mostly when the skew measure comes from the single stock options market (between 17 and 21 percent). As a robustness check, we note that the regression results are virtually unchanged by the usage of either *IV-sentiment 90-110* or *80-120* measures. As a first impression, these results imply that *IV-sentiment* is strongly connected to *DGspread* and to overweight of small probabilities.

Panel B shows that when we evaluate the multivariate regressions, we find that *IV-sentiment* is the most stable regressor with respect to coefficient signs, being negatively linked to *DGspread* across all regressions, and is always statistically significant. These regressions have high explanatory power (ranging from 41 to 61 percent), especially when considering the daily frequency, thus, potentially containing more noise than lower frequency data. In the multivariate regression we use the *IV-sentiment* 90-110, while the (unreported) results using *IV-sentiment* 80-120 are qualitatively the same. Due to likely multicollinearity in this multivariate model, we believe that our univariate models are more insightful than the former.

These findings strongly suggest that *DGspread* co-moves with our *IV-sentiment* measure within the three-, six-, and twelve-month maturities. Hence, we feel comfortable to interpret *IV-sentiment* as a proxy for overweighting of small probabilities, such as *DGspread*.

5. Robustness tests

5.1. *IV-sentiment versus single-market IV-sentiment strategies*

In Section 3.1 we reported that the strategy with the highest correlations with the high-frequency *IV-sentiment* strategy is the IC strategy, which relies on information from both the single stock and index option markets. Nevertheless, the correlation of the IV Skew 90 (minus ATM IV) strategy with the *IV-sentiment* one is also relatively high. Therefore, we investigate here how our *IV-sentiment* strategy differs from *single-market* strategies when information from both OTM puts and OTM calls are used, instead of the ATM IV. We investigate this to further understand the value added by the two separate aspects of the *IV-sentiment*, i.e. the usage of OTM IV only and the usage of joint information from the two option markets.

We begin by producing high-frequency strategies for four *Single-market* skew measures, namely: the *IV-Sentiment-Single 90-110*, the *IV-Sentiment-Single 80-120*, the *IV-Sentiment-Index 90-110* and the *IV-Sentiment-Index 80-120*, where:

$$IV\text{-}Sentiment\text{-}Single = OTMsinglestockput\,IV_{\tau p}\text{-}OTMsinglestockcallIV_{\tau c}, \quad (20)$$

$$IV\text{-}Sentiment\text{-}Index = OTMindexputIV_{\tau p}\text{-}OTMindexcallIV_{\tau c}, \quad (21)$$

which we refer to, in short, as *IVSentSingle* and *IVSentIndex*. These strategies have the exact same construction as the high-frequency *IV-sentiment* strategy described in Section 3.1. The four strategies are compared to the *IV-sentiment* using the same analytics as employed in Section 3.1, in specific, a risk-return comparison, a correlation matrix and tail-dependence measures.‡

In general, *IVSentSingle* and *IVSentIndex* strategies perform poorer than the *IV-sentiment* strategy. Both *IVSentSingle*

† The regression results reported here use RND kurtosis and skewness from index options ($m = io$). The results when RND is extracted from single stock options ($m = sso$) are unreported but qualitatively the same as the coefficient signs are equal to the reported ones, and regressions' explanatory power are roughly in the same range.

‡ Complete analytics comparing *IV-sentiment* with *IVSentSingle* and *IVSentIndex* strategies are reported in table 8 available in the online Appendix C.2, available at https://github.com/luizfelix/IV-Sentiment.

Table 7. Regression results: *Delta minus Gamma spread* and risk-neutral measures.

Panel A – Univariate regressions

Maturity	3m	6m	12m	3m	6m	12m	3m	6m	12m	3m	6m	12m
Intercept	0.019**	– 0.219***	– 0.436***	– 0.045***	– 0.254***	– 0.460***	– 0.295***	– 0.499***	– 0.683***	– 0.186***	– 0.368***	– 0.593***
	(0.007)	(0.010)	(0.009)	(0.006)	(0.008)	(0.007)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Skewness	0.122***	0.073***	0.054***									
	(0.003)	(0.004)	(0.003)									
Kurtosis				– 0.015***	– 0.009***	– 0.007***						
				(0.000)	(0.000)	(0.000)						
IV-sentiment 90-110							– 1.998***	– 2.774***	– 2.359***			
							(0.046)	(0.064)	(0.062)			
IV-sentiment 80-120										– 1.606***	– 2.438***	– 2.124***
										(0.042)	(0.058)	(0.056)
R^2	32%	9%	7%	30%	10%	7%	34%	46%	36%	30%	45%	35%
F-stats	1861.1	408.6	285.7	1714.6	423.4	315.6	2085.1	3423.7	2228.0	1707.2	3199.9	2121.2
AIC	– 2001	– 13	– 944	– 1900	– 27	– 972	– 2151	– 2093	– 2437	– 1895	– 1971	– 2368
BIC	– 1989	– 1	– 931	– 1888	– 14	– 959	– 2139	– 2081	– 2424	– 1883	– 1959	– 2355

Panel A – Univariate regressions (continuation)							Panel B - Multivariate regressions		
Maturity	3m	6m	12m	3m	6m	12m	3m	6m	12m
Intercept	– 0.195***	– 0.141***	– 0.332***	0.029	– 0.052	– 0.407***	– 0.273***	– 0.465***	– 0.495***
	(0.010)	(0.013)	(0.011)	(0.028)	(0.034)	(0.027)	(0.025)	(0.038)	(0.040)
Skewness							0.093***	0.000	– 0.032***
							(0.008)	(0.009)	(0.009)
Kurtosis							– 0.002**	– 0.007***	– 0.009***
							(0.001)	(0.001)	(0.001)
IV-sentiment 90-110							– 1.989***	– 2.462***	– 1.677***
							(0.063)	(0.108)	(0.126)
IV 110-ATM skew	1.082**	13.681***	16.172***				0.511	5.525***	8.106***
	(0.435)	(0.717)	(0.711)				(0.371)	(0.672)	(1.004)
IV 90-ATM skew				– 4.941***	– 8.399***	– 4.997***	3.876***	4.129***	0.000
				(0.557)	(0.903)	(0.993)	(0.391)	(0.732)	(0.933)
R^2	0%	17%	21%	4%	4%	1%	61%	53%	42%
F-stats	10.3	810.0	1065.3	148.4	177.3	49.4	1214.5	903.8	707.4
AIC	– 485	– 362	– 1612	– 621	202	– 717	– 4154	– 2636	– 3008
BIC	– 472	– 349	– 1599	– 608	215	– 705	– 4116	– 2598	– 2970

Panel A reports the regression results for equations (15)–(18) in an univariate setting. The dependent variable for these regressions is *Delta minus Gamma spread* (δ - γ), a proxy for overweight of small probabilities. As explanatory variables we specify the risk-neutral skewness and kurtosis, IV 110-ATM skew (from single stock options), IV 90-ATM skew (from index options), and our *IV-sentiment* measure in two permutations per maturity: (1) *IV-sentiment 90-110*, and (2) *IV-sentiment 80-120*. Our *IV-sentiment* measure is an IV skew measure that combines information from the index option market and the single stock option market, see equation (1). For instance, the *IV-sentiment 90-110* measure combines the IV from the 90 percent moneyness level from the index option market and the 110 percent moneyness level from the single stock option market. Panel B reports the regression results for equation (19) in a multivariate setting, in which *Delta minus Gamma spread* is regressed on the same set of explanatory variables. We report Newey-West adjusted standard errors in brackets. Asterisks ***, **, and * indicate significance at the one, five, and ten percent level, respectively.

strategies deliver negative returns within our sample. The *IVSentIndex* strategies show positive returns but they generate at best half of the performance delivered by the *IV-sentiment* strategy. Skewness is only positive for *IV-sentiment* strategy and negative for all *Single-market* strategies. Maximum drawdown is also the lowest for the *IV-sentiment* strategy. In line with these results and with our earlier analysis that split the performance of *IV-sentiment* strategy into the long and short legs (see Section 3.1), we find that the *IV-sentiment* strategy is more correlated with the *IVSentIndex* strategies (correlation between 0.45 and 0.50) than with *IVSentSingle* strategies (correlation between 0.28 and 0.32). Tail dependence with the *IV-sentiment* strategy is mostly higher for the *IVSentIndex* strategies as well and reaches a maximum of 56 percent for the conditional co-crash probability.

These results suggest, despite that the *IVSentIndex* strategies being more connected to the *IV-sentiment* one than to the *IVSentSingle* ones, that these strategies are still quite different from each other. This finding also suggests that the usage of joint information from both option markets adds value to IV-based sentiment strategies, and that it carries different information compared to *single market* strategies.

Subsequently, we run equations 6c and 6d with the inclusion of the *IVSentSingle* and *IVSentIndex* strategies as additional explanatory variables. The idea here is to better explain the *IV-sentiment* strategy and to improve our insights on what matters in predicting the equity risk premium. We find that *IVSentSingle* and *IVSentIndex* are both highly significant within the regressions, used either in isolation or jointly. Apart from this result, the inclusion of these variables more than doubles the explanatory power of the regressions, especially when *IVSentIndex* is added, indicating that *single-market* IV skew strategies are more strongly connected to the *IV-sentiment* strategy than traditional equity factors. For monthly regressions, the α that was positive and significant when *IVSentSingle* and *IVSentIndex* were not included in the regression becomes non-significant, indicating that these new factors can explain the skill of the *IV-sentiment* strategy. Still, the fact that the explanatory power of these regression are, at best, 26 percent for daily models and 49 percent for monthly models shows that the amount of unexplained variance of the *IV-sentiment* strategy remains large. This conclusion holds even after information from both option markets, in the form of separate *single-market* skews, is used, suggesting that the IV of OTM single stock puts and OTM index calls might contain more noise than information for predicting the equity risk premium.†

5.2. *Controlling for investors' optimism*

So far we have taken a myopic view on the root causes of overweighting of small probabilities by investors. For ease of exposition, we have assumed that overweight of tails is linked to preferences (i.e. a behavioral bias). However, we acknowledge that it is yet unclear whether the overweighting of small probabilities is caused solely by preferences or rather by biased beliefs (i.e. investors' expectations). Barberis (2013) eloquently discusses how both phenomena are distinctly different and how both (individually or jointly) may potentially explain the existence of overpriced OTM options. Therefore, in this section, we provide evidence that investor sentiment is linked to time-varying preferences for lottery tickets even after controlling for time-varying investor optimism. In other words, we reiterate the view taken in our paper that overweighting of tail events during periods of high sentiment reflects a behavioral bias rather than change in beliefs only.

We model investor optimism as proposed by Lemmon and Portniaguina (2006), who argue that investors' optimism is captured by consumer confidence. However, because consumer confidence may also reflect investor beliefs related to economic fundamentals, they propose that a purer measure of investors' optimism is the residuals of a regression between consumer confidence and a set of economic fundamentals:

$$CConf_t = \alpha_t[\tau] + EcoFund_t + \varepsilon[1]_t, \tag{22}$$

where, $\varepsilon[1]_t$ is the Michigan consumer confidence index (*CConf*) controlled for the effect of economic fundamentals, which is proxied here by the *Nowcasting* index of Beber *et al.* (2015). $\varepsilon[1]_t$ being then the investor optimism measure ($InvOpt_t$). With *CConf* and $InvOpt_t$ at hand, we run equations (23) and (24) below to filter out the effect of investor optimism (i.e. bias in beliefs) contained in *DGspreads*:

$$DGspread[\tau] = \alpha_t[\tau] + CConf_t + \varepsilon[2]_t, \tag{23}$$

$$DGspread[\tau] = \alpha_t[\tau] + InvOpt_t + \varepsilon[3]_t. \tag{24}$$

The obtained measures of overweighting of small probabilities solely linked to skewness preferences (i.e. a behavioral bias) are given by $\varepsilon[2]_t$ and $\varepsilon[3]_t$. We call $\varepsilon[2]_t$ the *residual DGspreads over consumer confidence* (*DGspread-CC*), and we call $\varepsilon[3]_t$ the *residual DGspreads over consumer confidence and fundamentals* (*DGspread-CCF*).

As a next step, we are interested to learn whether these purer measures of investor preferences are linked to sentiment. To do that we run a version of equation (13), where the explained variable is not *DGspread* but *DGspread-CC* and *DGspread-CCF*, which we generically call *residual DGspreads* (*ResDGSpread*) as in equation (25):

$$\begin{aligned} ResDGSpread[\tau]_t = {} & c + SENT_t + IISENT_t + E12_t + B/M_t \\ & + NTIS_t + TBL_t + INFL_t + CORPR_t + SVAR_t \\ & + CSP_t + \epsilon_t. \end{aligned} \tag{25}$$

Unreported results provides evidence that the residual version of *DGspreads* remains positively and statistically linked to sentiment as in equation (13). This is the case for both residual measures of *DGspreads* in multivariate and univariate settings. This result reiterates the interpretation supported through our paper that overweighting of small probabilities is caused, to some extent, by a behavioral bias. Further, the explanatory power of these regressions are smaller (at maximum 36 percent for multivariate models and 17 percent for univariate models) than the ones achieved by equation (13),

† Full results for equations 6c and 6d with the usage of *IVSentSingle* and *IVSentIndex* as additional explanatory variables are reported in table 9 of the online Appendix C.2, available at https://github.com/luizfelix/IV-Sentiment.

which reaches 57 percent for multivariate models and 32 for univariate models (reported by table 6), suggesting that *DGspreads* are partially driven by investor optimism.

In addition, we re-run equation (19) by replacing the explained variable *DGspreads* by our new measures *ResDGSpread*. The goal of this model, given by equation (26), is to check whether the residual versions of *DGspreads* remain linked to *IV-sentiment*:

$$ResDGSpread[\tau]_t = \alpha_t + SKEW_t^m(\tau) + KURT_t^m(\tau) + IVSent_t + \epsilon_t. \quad (26)$$

Unreported results indicate that *DGspreads* controlled by investor expectations (bias in beliefs) remain statistically and negatively linked to our proposed *IV-sentiment* indicator, as previous reported by table 7.† Similarly to our previous example, we find that the power of *IV-sentiment* in explaining *DGspreads*, ranging from 16 to 53 percent, is higher than in explaining *ResDGSpread*, ranging from 11 to 35 percent. The gap in explanatory power between the regressions that have the two types of *ResDGSpread* as explained variable is not substantial, suggesting that the adjustment made for economic conditions does impact results materially. These results suggest that *IV-sentiment* captures mostly investors' preference for skewness beyond capturing investors' optimism.

5.3. *Implied correlation and correlation risk premium factors*

As earlier suggested, the closest measure to our *IV-sentiment* indexes are the implied correlation (IC) and the correlation risk premia (CRP) measures of Driessen *et al.* (2013) and Buss *et al.* (2017). This is no coincidence, as all these measures are jointly calculated from the two types of equity option markets available, the index and the single stock option markets. Nevertheless, because both IC and CRP are computed from the entire cross-section of strikes and *IV-sentiment* is estimated from OTM options only, these measures are inherently different. In this robustness test we attempt to better understand the difference between these three measures.

First, we estimate a correlation matrix of the Buss *et al.* (2017) factors,‡ *IV-sentiment* measures and *DGspreads* using our full sample, see figure 7 available at our online Appendix. Buss *et al.* (2017) compute ICs for the standard maturities of 30, 91, 182, 273, and 365 days. CRPs are computed as IC minus the realized correlation from daily returns from the historical window equal to the maturity of the options used for a given IC.§ The results are as expected, to the extent that *IV-sentiment* is mostly correlated to IC (between 0.66 and 0.83) and *DGspreads* (0.72) rather than to the CRP (between 0.12 and 0.2). Additionally, we observe that IC is very weakly linked to *DGspreads* (between 0 and 0.2). This finding reiterates our opinion that IC seems not to reflect overweighting of small probabilities. The intuitive explanation for this result is the fact that IC is not locally linked to the tails of the risk neutral density implied by options but rather to the full cross-sectional of strikes, which largely overstates the information contained in close to at-the-money options versus OTM options.

Second, we are also interested in explaining the difference between IC and *IV-sentiment*. Therefore, we design equation (27), which has the spread IC_t-$IVSent_t$ as its explained variable:

$$IC_t[\tau] - IVSent_t[\tau] = \alpha_t[\tau] + SKEW_t^m(\tau) + KURT_t^m(\tau) + \epsilon_t. \quad (27)$$

The results suggest that the risk-neutral skweness and kurtosis explain a substantial part of the spread between IC_t and $IVSent_t$. The skewness and kurtosis factors are, not only highly significant, but regressions deliver R^2s between 49 and 57 percent.¶

Our interpretation of this result is that, as the 3rd and 4th moments of the implied risk-neutral distribution have a connection with the distribution tails, whereas the IC correlation does not, these two factors can to a large extent explain the difference between IC_t and $IVSent_t$.

6. Conclusion

End-users of OTM options tend to overweight tail events. This is a well accepted assumption, which applies to both OTM index puts and single stock calls, due to individual and institutional investors trading activity, respectively. Individual investors typically buy OTM single stock calls ('lottery tickets') to speculate on the upside of equities (indicating bullish sentiment), whereas institutional investors typically buy OTM index puts (portfolio insurance) to protect their large equity holdings (indicating bearish sentiment). Thus, we conjecture that information from options might capture those two opposite investors' risk attitudes and we propose a novel sentiment proxy: *IV-sentiment*.

The uniqueness of our *IV-sentiment* measure is that it is jointly calculated from the IV of index puts and single stock call options. In contrast with other measures jointly calculated from the index and single stock option markets, such as the implied correlation calculated from the entire cross-section of strikes, our measure uses OTM options only.

We find that our contrarian-trading strategies using our *IV-sentiment* measure produce economically significant risk-adjusted returns, reiterating earlier findings by the literature that equity markets are slow to incorporate the information embedded in implied volatility skews. The joint use of information from the single stock and index option markets seems to be the reason for the superior forecast ability of our *IV-sentiment* measure, because factors that use implied volatility skews from a single market achieve significantly inferior

† Full results for this Section are provided by tables 10 and 11 reported in the online Appendix C.3, available at https://github.com/luizfelix/IV-Sentiment.

‡ Data use in Buss *et al.* (2017) is kindly provided by Grigory Vilkov at http://www.vilkov.net/codedata.html.

§ The estimated correlation matrix is shown in figure 7, which is reported in the online Appendix C.4 (available at https://github.com/luizfelix/IV-Sentiment).

¶ Complete results of the estimation of equation (27) are reported in table 12 contained in the online Appendix C.4, available at https://github.com/luizfelix/IV-Sentiment.

results. The performance of our *IV-sentiment* measure seems also more consistent in delivering a positive information ratio than the Baker and Wurgler (2007) sentiment factor, beyond being more positively skewed, having a shorter horizon and allowing for daily rebalancing.

Our *IV-sentiment* factor seems to forecast returns as well as other well-known predictors of equity returns. Since it is uncorrelated to these predictors of the equity risk-premium, it significantly improves the quality of predictive models, especially by the usage of constrained ensemble models and vis-à-vis unregularized machine learning techniques. Finally, our factor has limited exposure to a set cross-sectional equity factors and seems valuable in preventing momentum crashes.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Ait-Sahalia, Y. and Lo, A., Nonparametric risk management and implied risk aversion. *J. Econom.*, 2000, **94**, 90–51.

Amin, K., Coval, J. and Seyhun, H., Index option prices and stock market momentum. *J. Bus.*, 2004, **77**, 835–874.

Ang, A., Chen, J. and Xing, Y., Risk, return and dividends. *Rev. Financ. Stud.*, 2006, **19**, 1191–1239.

Ang, A. and Liu, J., Risk, return and dividends. *J. Financ. Econ.*, 2007, **85**, 1–38.

Baker, M. and Wurgler, J., Investor sentiment in the stock market. *J. Econ. Perspect.*, 2007, **21**, 129–151.

Bakshi, G., Kapadia, N. and Madan, D., Stock returns characteristics, skew laws, and the differential pricing of individual equity options. *Rev. Financ. Stud.*, 2003, **16**, 101–143.

Barberis, N., The psychology of tail events: Progress and challenges. *Am. Econ. Rev.*, 2013, **103**, 611–616.

Barberis, N. and Huang, M., Stocks as lotteries: The implication of probability weighting for security prices. *Am. Econ. Rev.*, 2008, **98**, 2066–2100.

Bates, D., The crash of '87: Was it expected? The evidence from options markets. *J. Finance*, 1991, **46**, 1009–1044.

Bates, D., Empirical option pricing: A retrospection. *J. Econom.*, 2003, **116**, 387–404.

Beber, A., Brandt, M. and Luisi, M., Distilling the macroeconomic news flow. *J. Financ. Econ.*, 2015, **117**, 489–507.

Bliss, R. and Panigirtzoglou, N., Option-implied risk aversion estimates. *J. Finance*, 2004, **59**, 407–446.

Bollen, N. and Whaley, R., Does net buying pressure affect the shape of implied volatility function? *J. Finance*, 2004, **59**, 711–753.

Bollerslev, T., Tauchen, G. and Zhou, H., Expected stock returns and variance risk premia. *Rev. Financ. Stud.*, 2009, **22**, 4463–4492.

Boyer, B. and Vorkink, K., Stock options as lotteries. *J. Finance*, 2014, **69**, 1485–1527.

Brunnermeier, M. and Pedersen, L., Market liquidity and funding liquidity. *Rev. Financ. Stud.*, 2009, **22**, 2201–2238.

Buss, A., Schoenleber, L. and Vilkov, G., Expected stock returns and the correlation risk premium. *CEPR Discussion Paper No. DP12760*, pp. 1–61, 2017.

Campbell, J. and Cochrane, J., By force of habit: A consumption-based explanation of aggregate stock market behavior. *J. Polit. Econ.*, 1999, **107**, 205–251.

Campbell, J. and Thompson, S., Predicting the equity premium out of sample: Can anything beat the historical average?. *Rev. Financ. Stud.*, 2008, **21**, 1509–1531.

Carhart, M., On persistence in mutual fund performance. *J. Finance*, 1997, **52**, 57–82.

Chan, L., Chen, H.L. and Lakonishok, J., On mutual fund investment styles. *Rev. Financ. Stud.*, 2002, **15**, 1407–1437.

Cornell, B., The pricing of volatility and skewness: A new interpretation. *J. Investing*, 2009, **18**, 27–30.

De Long, J., Shleifer, A., Summers, L. and Waldmann, R., Noise trader risk in financial markets. *J. Polit. Econ.*, 1990, **98**, 703–738.

Dierkes, M., Option-implied risk attitude under rank-dependent utility. Unpublished working paper. University of Münster, Münster, Germany, 2009.

Doran, J., Peterson, D. and Tarrant, B., Is there information in the volatility skew? *J. Future Markets*, 2007, **27**, 921–959.

Driessen, J., Maenhout, P. and Vilkov, G., Option-implied correlations and the price of correlation risk. SSRN working paper no. 2166829, pp. 1–46, 2013.

Fama, E. and French, K., The cross-section of expected stock returns. *J. Finance*, 1992, **47**, 427–465.

Fama, E. and French, K., A five-factor asset pricing model. *J. Financ. Econ.*, 2015, **116**, 1–22.

Fama, E. and French, K., Dissecting anomalies with a five-factor model. *Rev. Financ. Stud.*, 2016, **29**, 69–103.

Félix, L., Kräussl, R. and Stork, P., Single stock call options as lottery tickets. *J. Behav. Financ.*, 2019, **20**, 385–407.

Figlewski, S., Estimating the implied risk neutral density for the us market portfolio. In *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, edited by Tim Bollerslev, Jeffrey Russell, and Mark Watson, 2010 (Oxford University Press: Oxford).

Frazzini, A. and Pedersen, L., Betting against beta. *J. Financ. Econ.*, 2014, **111**, 1–25.

Frijns, B., Huynh, T., Tourani-Rad, A. and Westerholm, P., Institutional trading and asset pricing. *J. Bank. Financ.*, 2018, **89**, 59–77.

Garleanu, N., Pedersen, L.H. and Poteshman, A.M., Demand-based option pricing. *Rev. Financ. Stud.*, 2009, **22**, 4259–4299.

Han, B., Investor sentiment and option prices. *Rev. Financ. Stud.*, 2008, **21**, 387–414.

Hartmann, P., Straetmans, S. and de Vries, C., Asset market linkages in crisis periods. *Rev. Econ. Stat.*, 2004, **86**, 313–326.

Harvey, C. and Siddique, A., Conditional skewness in asset pricing tests. *J. Finance*, 2000, **55**, 1263–1295.

Hastie, T., Tibshirani, R. and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2008 (Springer-Verlag: New York).

Jackwerth, J. and Rubinstein, M., Recovering probability distributions from option prices. *J. Finance*, 1996, **51**, 1611–1631.

Kent, D. and Moskowitz, T., Momentum crashes. *J. Financ. Econ.*, 2016, **122**, 221–247.

Kliger, D. and Levy, O., Theories of choice under risk: Insights from financial markets. *J. Econ. Behav. Organ.*, 2009, **71**, 330–346.

Krishnan, C.N.V., Petkova, R. and Ritchken, P., Correlation risk. *J. Empir. Finance*, 2009, **16**, 353–367.

Lakonishok, J., Lee, I., Pearson, N. and Poteshman, A., Option market activity. *Rev. Financ. Stud.*, 2007, **20**, 813–857.

Lemmon, M. and Ni, S., The effects of investor sentiment on speculative trading and prices of stock and index options. SSRN working paper no. 1572427, 2011.

Lemmon, M. and Portniaguina, E., Consumer confidence and asset prices: Some empirical evidence. *Rev. Financ. Stud.*, 2006, **19**, 1499–1529.

Li, J., Yao, Y., Chen, Y. and Lee, C., Option prices and stock market momentum: Evidence from China. *Quant. Finance*, 2018, **18**, 1517–1529.

Longstaff, F., Option pricing and the martingale restriction. *Rev. Financ. Stud.*, 1995, **8**, 1091–1124.

Mahani, R. and Poteshman, A., Overreaction to stock market news and misevaluation of stock prices by unsophisticated investors: Evidence from the options market. *J. Empir. Finance*, 2008, **15**, 635–655.

Mitton, T. and Vorkink, K., Equilibrium underdiversification and the preference for skewness. *Rev. Financ. Stud.*, 2007, **20**, 1255–1288.

Moskowitz, T., Ooi, Y.H. and Pedersen, L.H., Time series momentum. *J. Financ. Econ.*, 2012, **104**, 228–250.

Polkovnichenko, V. and Zhao, F., Probability weighting functions implied in option prices. *J. Financ. Econ.*, 2013, **107**, 580–609.

Pollet, J. and Wilson, M., Average correlation and stock market returns. *J. Financ. Econ.*, 2008, **96**, 364–380.

Prelec, D., The probability weighting function. *Econometrica*, 1998, **66**, 497–527.

Rapach, D., Strauss, J. and Zhou, G., Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Rev. Financ. Stud.*, 2010, **23**, 821–862.

Rosenberg, J. and Engle, R., Empirical pricing kernels. *J. Financ. Econ.*, 2002, **64**, 341–372.

Rubbaniy, G., Asmerom, R., Rizvi, S. and Naqvi, B., Do fear indices help predict stock returns?. *Quant. Finance*, 2014, **14**, 831–847.

Tversky, A. and Kahneman, D., Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk. Uncertain.*, 1992, **5**, 297–323.

Vilkov, G. and Xiao, Y., Option-implied information and predictability of extreme returns. SAFE (Goethe University Frankfurt) Working Paper Series 5, 1–36, 2013.

Welch, I. and Goyal, A., A comprehensive look at the empirical performance of equity premium prediction. *Rev. Financ. Stud.*, 2008, **21**, 1455–1508.

Xing, Y., Zhang, X. and Zhao, R., What does the individual option volatility smirk tell us about future equity returns? *J. Financ. Quant. Anal.*, 2010, **45**, 641–662.

Appendix 1. Methodology

A.1. Subject density function estimation

We hereby present the derivations required to achieve equation (9) in the main text, equation (A7) here, from equation (8), called here equation (A1):

$$\frac{f_Q(S_T)}{w'(F_P(S_T)) \cdot f_P(S_T)} = \varsigma(S_T), \tag{A1}$$

where $f_P(S_T)$ is the 'real-world' probability distribution, $f_Q(S_T)$ is the RND, $\varsigma(S_T)$ is the pricing kernel, w is the weighting function, and $F_P(S_T)$ is the 'real-world' cumulative density function.

The first step of our derivation entails re-arranging equation (A1) into (A2b) via equation (A2a), which demonstrates that for the CPT to hold, the subjective density function should be consistent with the probability weighted EDF:

$$\underbrace{f_Q(S_T)}_{\text{RND}} = \underbrace{w'(F_P(S_T))}_{\text{probability weighing}} \cdot \underbrace{f_P(S_T)}_{\text{EDF}} \cdot \underbrace{\varsigma(S_T)}_{\text{pricing kernel}}, \tag{A2a}$$

$$\underbrace{f_Q(S_T)}_{\text{RND}} = \underbrace{f_{\tilde{P}}(S_T)}_{\text{probability weighted EDF}} \cdot \underbrace{\varsigma(S_T)}_{\text{pricing kernel}}, \tag{A2b}$$

$$\underbrace{\frac{f_Q(S_T)}{\lambda \frac{U'(S_T)}{U'(S_t)}} = \frac{f_Q(S_T)}{\varsigma(S_T)}}_{\text{Subjective density}} = \underbrace{f_{\tilde{P}}(S_T)}_{\text{probability weighted EDF}}. \tag{A3}$$

Following Ait-Sahalia and Lo (2000) and Bliss and Panigirtzoglou (2004), equation (A3) can be manipulated so that the time-preference constant Λ of the pricing kernel vanishes, producing equation (A4), which directly relates the probability weighted EDF, the RND, and the marginal utility, $U'(S_T)$:

$$\underbrace{f_{\tilde{P}}(S_T)}_{\text{probability weighted EDF}} = \underbrace{\frac{\lambda \frac{U'(S_T)}{U'(S_t)} Q(S_T)}{\int \frac{U'(S_t)}{U'(x)} Q(x)\,\mathrm{d}x} = \frac{\frac{f_Q(S_T)}{U'(S_T)}}{\int \frac{f_Q(x)}{U'(x)}\mathrm{d}x}}_{\text{Generic subjective density function}}, \tag{A4}$$

where $\int (Q(x)/U'(x))\,\mathrm{d}x$ normalizes the resulting subjective density function to integrate to one. Once the utility function is estimated, equation (A4) allows us to convert RND into the probability weighted EDF. equation (A4) can also be used to estimate the subjective density function for an (rational) investor that has a power or exponential utility function, by disregarding the weighting function $W(\cdot)$, so the LHS of the equation becomes $f_p(S_T)$. In the remainder of the paper we call these subjective distributions power and exponential density functions. As we hypothesize that the representative investor has a CPT utility function, its marginal utility function is $U'(S_T) = \upsilon'(S_T)$, and, thus, $\upsilon'(S_T) = \alpha S_T^{\alpha-1}$ for $S_T >= 0$, and $\upsilon'(S_T) = -\lambda\beta(-S_T)^{\beta-1}$ for $S_T < 0$, leading to equation (A5):

$$f_{\tilde{P}}(S_T) = \frac{\frac{f_Q(S_T)}{\alpha S_T^{\alpha-1}}}{\int \frac{f_Q(x)}{\alpha x^{\alpha-1}}\mathrm{d}x} \quad \text{for } S_T \geq 0, \quad \text{and} \tag{A5}$$

$$\underbrace{f_{\tilde{P}}(S_T)}_{\text{probability weighted EDF}} = \underbrace{\frac{\frac{f_Q(S_T)}{-\lambda\beta(-S_T)^{\beta-1}}}{\int \frac{f_Q(x)}{-\lambda\beta(-x)^{\beta-1}}\mathrm{d}x}}_{\text{Partial CPT density function}} \quad \text{for } S_T < 0, \tag{A6}$$

Equations (A5) and (A6), hence, relate the EDF where probabilities are weighted according to the CPT probability distortion functions, on the LHS, to the subjective density function derived from the CPT value function, on the RHS, separately for gains and losses, i.e. the PCPT density function. The relationships specified by equations (A5) and (A6) fully state the relation we would like to depict, although one additional manipulation is convenient for our argumentation. Assuming that the function $w(F_P(S_T))$ is strictly increasing over the domain [0,1], there is a one-to-one relationship between $w(F_P(S_T))$ and a unique inverse $w^{-1}(F_P(S_T))$. As such, the result $f_{\tilde{P}}(S_T) = w'(F_P(S_T))f_P(S_T)$ also implies $f_{\tilde{P}}(S_T).(w^{-1})'(F_P(S_T)) = f_P(S_T)$.† This outcome allows us to directly relate the original EDF to the CPT subjective density function, by 'undoing' the effect of the CPT probability distortion functions within the PCPT density function:

$$\underbrace{f_P(S_T)}_{\text{EDF}} = \underbrace{\frac{\frac{f_Q(S_T)}{\upsilon'(S_T)}}{\int \frac{f_Q(x)}{\upsilon'(x)}\mathrm{d}x}(w^{-1})'(F_P(S_T))}_{\text{CPT density function}}. \tag{A7}$$

Thus, once the relation between the probability weighting function of EDF and the PCPT density is established, as in equations (A5) and (A6), one can eliminate the weighting scheme affecting returns by applying the inverse of such weightings to the subjective density function without endangering such equalities, as in equation (A7), numbered equation (9) in the main text.

A.2. Weighted average single stock IV and implied correlation approximations

In the following, we derive the weighted average single stock IV, equation (A8l), and the implied correlation approximation, equation (A8j), as given by equation (7) in the main text:

$$\sigma_P^2 = \sum_{i=1}^{n} w_i^2\sigma_i^2 + \sum_{i \neq j}^{n} w_i w_j \rho_{ij}\sigma_i\sigma_j \tag{A8a}$$

† A drawback of the CPT model is that it allows for non-strictly increasing functions, which would not allow invertibility. This is the reason why the newer literature on probability distortions functions favors other strictly monotonic functions, such as Prelec's (1998) $w(p) = e^{-(-\ln(p))^\delta}$, as the weighting functions. Nevertheless, because the CPT parameters of our interest ($\gamma = 0.61$; $\delta = 0.69$) impose strict monotonicity, we can obtain the inverse of the probability function, $w^{-1}(p)$ numerically.

Starting from the portfolio variance σ_P^2 formula given by equation (A8a), where i and j are indexes for the portfolio constituents, this relation can be re-written for a equity index as:

$$\sigma_I^2 = \sum_{i,j=1}^{n} w_i w_j \rho_{ij} \sigma_i \sigma_j, \tag{A8b}$$

implying that

$$\sum_{i \neq j}^{n} w_i w_j \rho_{ij} \sigma_i \sigma_j = \sum_{i,j=1}^{n} w_i w_j \rho_{ij} \sigma_i \sigma_j - \sum_{i=1}^{n} w_i^2 \sigma_i^2 \tag{A8c}$$

where,

$$\rho_{ij}(x) = \begin{cases} \bar{\rho}, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases} \tag{A8d}$$

and where σ_I^2 is the equity index option-implied variance. Then, assuming $\bar{\rho}$ as the estimator for average stock correlation we have:

$$\sigma_I^2 = \bar{\rho} \sum_{i \neq j}^{n} w_i w_j \sigma_i \sigma_j + \sum_{i=1}^{n} w_i^2 \sigma_i^2, \tag{A8e}$$

which, given equality A8c, can be re-written as:

$$\sigma_I^2 = \bar{\rho} \sum_{i,j=1}^{n} w_i w_j \sigma_i \sigma_j - \bar{\rho} \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} w_i^2 \sigma_i^2, \tag{A8f}$$

$$= \bar{\rho} \left(\sum_{i=1}^{n} w_i \sigma_i \right)^2 - \bar{\rho} \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} w_i^2 \sigma_i^2, \tag{A8g}$$

$$= \bar{\rho} \left(\left(\sum_{i=1}^{n} w_i \sigma_i \right)^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2 \right) + \sum_{i=1}^{n} w_i^2 \sigma_i^2, \tag{A8h}$$

$$\bar{\rho} = \frac{\sigma_I^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2}{(\sum_{i=1}^{n} w_i \sigma_i)^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2}. \tag{A8i}$$

As $\sum_{i=1}^{n} w_i^2 \sigma_i^2$ is relatively small, we can simplify equation (A8i), the implied correlation, into the approximated implied correlation given by equation (A8j). Note that, as $\sum_{i=1}^{n} w_i^2 \sigma_i^2$ is always positive, the approximated implied correlation will always overstate the true implied correlation:

$$\bar{\rho} \approx \frac{\sigma_I^2}{(\sum_{i=1}^{n} w_i \sigma_i)^2}. \tag{A8j}$$

Further, in order to obtain the weighted average single stock implied volatility,equation (A8l), we square root both sides of the approximation and re-arrange their terms:

$$\sqrt{\bar{\rho}} \approx \frac{\sigma_I}{(\sum_{i=1}^{n} w_i \sigma_i)} \tag{A8k}$$

with

$$\sum_{i=1}^{n} w_i \sigma_i \approx \frac{\sigma_I}{\sqrt{\bar{\rho}}}. \tag{A8l}$$

We note that, given equality (A8c), equation (A8i) can be re-written as:

$$\bar{\rho} = \frac{\sigma_I^2 - \sum_{i,j=1}^{n} w_i^2 \sigma_i^2}{\sum_{i \neq j}^{n} w_i w_j \sigma_i \sigma_j} = \frac{\sigma_I^2 - \sum_{i=1}^{n} \sum_{j=1}^{n} w_i^2 \sigma_i^2}{\sum_{i=1}^{n} \sum_{i \neq j} w_i w_j \sigma_i \sigma_j}, \tag{A8m}$$

which is the implied correlation (IC) measure employed by Driessen *et al.* (2013).

A.3. *Conditional co-crash probabilities*

We use a bivariate Extreme Value Theory (EVT) method to calculate commonality on historical tail returns for the strategies highlighted in Section 3.1. EVT is well suited to measure contagion risk because it does not assume any specific return distribution. Our approach estimates how likely it is that one stock will experience a crash beyond a specific extreme negative return threshold conditional on another stock crash beyond an equally probable threshold. We refer to Hartmann *et al.* (2004) who use the conditional co-crash (CCC) probability estimator, which is applied to each pair of stocks in our sample, as follows:

$$\widehat{CCC}_{ij} = 2 - \frac{1}{k} \sum_{t=1}^{N} I[V_{it} > x_{i,N-k} \text{ or } V_{jt} > x_{j,N-k}], \tag{A9}$$

where the function I is the crash indicator function, in which $I = 1$ in case of a crash, and $I = 0$ otherwise, V_{it} and V_{jt} are returns for stocks i and j at time t; $x_{i,N-k}$, and $x_{j,N-k}$ are extreme crash thresholds. The estimation of the CCC-probabilities requires setting k as the number of observations used in equation (A9).

Appendix 2. Equity market control variables and predictors

The complete set and summarized descriptions of variables provided by Welch and Goyal (2008)† that are used in our study is given as:

(1) *Dividend price ratio (log), D/P*: Difference between the log of dividends paid on the S&P 500 index and the log of stock prices (S&P 500 index).
(2) *Dividend yield (log), D/Y* : Difference between the log of dividends and the log of lagged stock prices.
(3) *Earnings, E12*: 12-month moving sum of earnings on teh S&P500 index.
(4) *Earnings-price ratio (log), E/P*: Difference between the log of earnings on the S&P 500 index and the log of stock prices.
(5) *Dividend-payout ratio (log), D/E*: Difference between the log of dividends and the log of earnings.
(6) *Stock variance, SVAR*: Sum of squared daily returns on the S&P 500 index.
(7) *Book-to-market ratio, B/M*: Ratio of book value to market value for the Dow Jones Industrial Average.
(8) *Net equity expansion, NTIS*: Ratio of twelve-month moving sums of net issues by NYSE-listed stocks to total end-of-year market capitalization of NYSE stocks.
(9) *Treasury bill rate, TBL*: Interest rate on a three-month Treasury bill.
(10) *Long-term yield, LTY* : Long-term government bond yield.
(11) *Long-term return, LTR*: Return on long-term government bonds.
(12) *Term spread, TMS*: Difference between the long-term yield and the Treasury bill rate.
(13) *Default yield spread, DFY* : Difference between BAA- and AAA-rated corporate bond yields.
(14) *Default return spread, DFR*: Difference between returns of long-term corporate and government bonds.
(15) *Cross-sectional premium, CSP*: measures the relative valuation of high- and low-beta stocks.
(16) *Inflation, INFL*: Calculated from the CPI (all urban consumers) using $t-1$ information due to the publication lag of inflation numbers.

† Available at http://www.hec.unil.ch/agoyal/.